1987

The aggregation of population groups to improve the predictability of Marine Corps officer attrition estimation.

Larsen, Randall W.; Read, Robert R.

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

THE AGGREGATION OF POPULATION GROUPS
TO IMPROVE THE PREDICTABILITY OF
MARINE CORPS OFFICER ATTRITION ESTIMATION

by

Randall W. Larsen

December 1987

Thesis Advisor: Robert R. Read

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE

EPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
ECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
ECLASSIFICATION / DOWNGRADING SCHEDULE	
RFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
IAME OF PERFORMING ORGANIZATION val Postgraduate School	6b. OFFICE SYMBOL (If applicable) Code 54
	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
DDRESS (City, State, and ZIP Code) terey, California 93943-5000	7b. ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
IAME OF FUNDING / SPONSORING RGANIZATION	8b. OFFICE SYMBOL (If applicable)
	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

ITLE (Include Security Classification)
: AGGREGATION OF POPULATION GROUPS TO IMPROVE THE PREDICTABILITY OF MARINE RPS OFFICER ATTRITION ESTIMATION

ERSONAL AUTHOR(S)
rsen, Randall W.

TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
ter's Thesis	FROM _____ TO _____	1987, December	110

UPPLEMENTARY NOTATION

COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Loss Rate Cluster Analysis; Officer Attrition Rate

BSTRACT (Continue on reverse if necessary and identify by block number)

This thesis presents an algorithm for the aggregation of low inventory tegories (small cells) which characterize the population of Marine Corps, restricted, active duty officers. The basis for aggregating these small lls is the degree of homogeneity of historical attrition rates. The chniques of hierarchical cluster analysis are applied to the small cell oblem in lieu of existing functional and organizational structures.

This research demonstrates the adaptability of cluster analysis to loss te aggregation and provides a shell for more refined model applications. rther, statistical stability and attrition rate homogeneity have been troduced to allow for subsequent application of shrinkage type parameter

ISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
NAME OF RESPONSIBLE INDIVIDUAL of. Robert R. Read	22b. TELEPHONE (Include Area Code) (408) 646-2382
	22c. OFFICE SYMBOL Code 55Re

FORM 1473, 84 MAR 83 APR edition may be used until exhausted.
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

☆ U.S. Government Printing Office: 1986—606-24.

#19 - ABSTRACT - (CONTINUED)

estimation methods associated with the development of an officer attrition rate generator.

Approved for public release; distribution is unlimited

The Aggregation of Population Groups to Improve the Predictability of Marine Corps Attrition Estimation

by

Randall W. Larsen
Captain, United States Marine Corps
B.S., Iowa State University, 1976

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
December 1987

ABSTRACT

This thesis presents an algorithm for the aggregation of low inventory categories (small cells) which characterize the population of Marine Corps, unrestricted, active duty officers. The basis for aggregating these small cells is the degree of homogeneity of historical attrition rates. The techniques of hierarchical cluster analysis are applied to the small cell problem in lieu of existing functional and organizational structures.

This research demonstrates the adaptability of cluster analysis to loss rate aggregation and provides a shell for more refined model applications. Further, statistical stability and attrition rate homogeneity have been introduced to allow for subsequent application of shrinkage type parameter estimation methods associated with the development of an officer attrition rate generator.

THESIS DISCLAIMER

The reader is cautioned that computer programs developed in this research may not have been exercised for all cases of interest. While every effort has been made, within the time available, to ensure that the programs are free of computational and logic errors, they cannot be considered validated. Any application of these programs without additional verification is at the risk of the user.

TABLE OF CONTENTS

I. INTRODUCTION

A. GENERAL

The purpose of this research is to enhance the predictability of Marine Corps officer attrition estimation. This paper is in support of a large, on-going effort concerning manpower model development and system integration under the broad title of Officer Planning and Utilization System (OPUS). Defense Systems Associates, Inc. (DSAI), Rockville, Maryland, is the contracted system developer of OPUS. The Navy Personnel Research and Development Center (NPRDC), San Diego, California, aided by Professor R.R. Read, Naval Postgraduate School, Monterey, California, is developing an officer attrition rate generator integral to OPUS. This thesis is conducted in conjunction with the work of NPRDC and Professor Read.

B. BACKGROUND

The United States Marine Corps officer corps is a hierarchical force of approximately 20,000 men and women. Marine Corps officer manpower planners are tasked with forecasting accessions, losses and promotions in order to meet present and anticipated personnel demands.

In military manpower planning models, personnel flows are generally the result of vacancies created within the system. For the most part, vacancies are the result of

losses. Losses in the rank hierarchy prompt promotions. Vacancies also create needs for accessions to replenish desired total force levels. As promotions and accessions are directly associated with losses, both are dependent on accurate loss forecasting. Underestimating losses can result in too few accessions, too few promotions, and ultimately may affect mission readiness. Overestimating losses can lead to too many accessions, underutilization of personnel, delays in promotion, and potential cost overruns.

The manpower planners of the Marine Corps manage and organize officers based on rank and military occupational specialty (MOS). As such, losses must be anticipated for each rank category and MOS. In order to project comprehensively the effects of attrition on the total force structure, losses are categorized by type and several descriptive variables associated with officer attrition behavior. The definition and discussion of the loss types and descriptive classifications are provided in Section D of this chapter.

When the various loss categories and all the defined descriptive variables are considered simultaneously in a multidimensional array, the number of potential individual cells exceeds four billion. As the officer population barely exceeds 20,000, the vast majority of the cells are unoccupied for either structural or sampling reasons. An example of an unoccupied cell due to structural reasons

would be a cell identifying lieutenant colonels, of any particular specialty, with six years of commissioned service. Such officers do not exist. Structurally zero inventories may be considered permanent conditions.

An unoccupied cell described as a sampling zero occurs due to chance and is not necessarily a permanent condition. In such a case a particular rank, MOS, and YCS combination may not exist during a particular year. This condition may change the following year as a result of promotions, accumulating YCS, or change of MOS.

The situation of sparse data over a large number of cells makes the task of accurate empirical rate estimation difficult. Small populations of characteristically limited and sporadic data lead to statistical instability, which in turn aggravates the rate forecasting problem [Ref. 1:p. 13; Ref. 2:p. 10; Ref. 3:p. 2]. This situation has been referred to as the "small cell problem" [Ref. 2:p. 10].

Presently, a highly comprehensive modeling system is being developed to predict future states of the officer force structure. This system, OPUS, is a computer-based planning tool reliant on predicted loss rates and target strength requirements [Ref. 4:pp. 2-1--2-59]. The Marine Corps Officer Rate Projector (MCORP) is the source of loss rate forecasts [Ref. 5:pp. 1-1--1-6]. Within MCORP, a computer program algorithm provides an automated calculation to meet certain computational requirements of the loss rate

forecasting system and represents the current solution to the small cell problem.

C. RESEARCH QUESTION

The primary research question is how to aggregate low inventory, officer categories (small cells) into sets of homogeneous attrition behavior in order to enhance forecasting techniques of developing manpower planning models. The solution must be a dynamic scheme in which small cells are aggregated in response to user designated minimum inventory thresholds. More, the methodology must reflect a versatile and flexible nature adaptive to changing conditions and the needs of manpower planners. This research effort will group officers of similar rank, years of commissioned service, and occupational specialty, stressing similarity of historical loss rates.

Subsidiary research questions include first, what features constitute a small cell and which categories represent small cells? Secondly, which small cells exhibit similar loss rate behavior? Finally, how can the aggregation of small cells be accomplished in order to meet the needs of Marine Corps planners and developing manpower models?

D. KEY TERMS

The terms loss and attrition will be used interchangeably. Losses and loss rates describe the flow of

officers from particular cells characterized by MOS, YCS, rank, etc. Flows may be from one cell to another within the Marine Corps or from a cell directly to the civilian labor market. Flows within the Marine Corps represent a loss only to the former cell not to the Service. Movement due to promotion, accumulation of service time, or changes of MOS are examples. Officers exiting a cell to the civilian labor force constitute an inventory loss to the Marine Corps. This project will focus on the attrition and the attrition rates of those leaving the Service.

The following terms will be used frequently in this analysis within the narrow context of Marine Corps officer manpower management:

- Accession--Accession refers to the commissioning of a new officer into the Marine Corps.
- Attrition--Attrition is the loss of an officer from the Service.
- Failed Select--Failed Select describes an officer not selected for promotion from either within or above the promotion zone. A lieutenant or captain who twice fails to be selected for promotion to a fixed rank must leave the Service. A major, lieutenant colonel, or colonel who twice fails to be selected for promotion to the next rank is limited to active service of 20, 26, or 30 years respectively.
- MCORP (Marine Corps Officer Rate Projector)--MCORP is an interactive software system which calculates Marine Corps officer loss rates based on historical attrition data.
- LOS (Length of Service)--LOS refers to the cumulative number of years served since date of service entry.
- MOS (Military Occupational Specialty)--MOS is a four-digit code identifying specific, skill-related classifications of Marines.

- OPUS (Officer Planning and Utilization System)--OPUS is a set of comprehensive computer-based models designed to support the data processing and forecasting requirements of Marine Corps officer planners.

- Regular Officer--A regular officer is an officer designated for long-term active duty, whose Service longevity is limited only by continued promotion and the statutory limits of service.

- Reserve Officer--A reserve officer is an officer designated to a fixed length of service. Such an officer may or may not be on active duty.

- YCS (Years of Commissioned Service)--YCS refers to the cumulative number of years served since date of commissioning.

In Table 1 is found a general description of the existing officer classification system and the extent of the classification alternatives. Appendix A offers a detailed explanation of all classifications within the data format.

E. SCOPE OF THE THESIS

For the stated purpose of this project, research will be limited to active duty, unrestricted, officers from the rank of second lieutenant to colonel. The management of inactive duty officers (inactive duty reservists and retirees) is sufficiently different to be excluded from OPUS and therefore of little relevance to this study. Limited duty officers (LDOs) are addressed separately in OPUS due to their unique career paths and characteristics of service; thus, this category will not be included in this thesis. Finally, the Marine Corps general officers (flag-rank) and warrant officers will not be discussed in this research project. General officers are an extremely small component

TABLE 1

OFFICER CLASSIFICATIONS

Designation	Description	Alternatives
LOSS	Retirement, Release, Discharge, Resignation, and Other.	5
RANK	Warrant Officer to Colonel. Differentiated further as to restricted or unrestricted and Failed-Select or Nonfailed Select.	21
MOS	Military Occupational Specialty. This category may be of further expanded by considering Secondary and Additional MOSs.	140+
YCS	Years of Commissioned Service. One to 31 plus.	31
SE	Source of Entry	14
SC	Service Component: Regular, Reserve to Regular, and Reserve Service	3
SS	Service School Completion	7
SEX	Sex	2
RACE	White, Black, Hispanic and Other	4
EDUC	Educational Attainment: Non-college grad, Four-year degree, Masters Degree, and Doctorate	4

of the total force structure, with required management taking place at the highest Service level. The warrant officers represent a narrowly defined population associated

with limited MOSs linked to the LDO categories and, as a group, have exhibited strong statistical stability in attrition behavior.

F. ORGANIZATION

In Chapter II a synopsis is presented of previous research pertinent to this thesis. In Chapter II is also provided a brief review of the theoretical and operational literature relevant to the research effort.

The structure and content of the utilized data bases are explained in Chapter III.

Chapter IV is begun with an explanation of the existing methodology for small cell aggregation within MCORP. The rationale is then given for the selection of cluster analysis in solving the small cell problem. Finally, the concepts and characteristics of the chosen technique are detailed.

The discussion in Chapter V describes the specific application of the clustering procedure to the research problem as well as the validation and analysis of the results.

Chapter VI presents the thesis summary and recommendations for ultimate application and maintenance of the improved methodology in the Marine Corps manpower planning system.

The appendices contain various details of interest to the reader desiring a more thorough explanation or

background on data format, applied computer programs, related cluster criterion testing, and other supporting material.

II. LITERATURE REVIEW

A. PRIOR STUDIES

This project should be recognized as a logical continuation of recent work done by Majors D.D. Tucker, USMC, and J.R. Robinson, USMC, and Colonel Amin Elseramegy, Egyptian Air Force, in their separate theses, at the Naval Postgraduate School.

In his September 1985 thesis, Tucker [Ref. 1] demonstrated the application of statistical shrinkage type parameter estimation techniques to the problem of small cells. His results were promising, though exploratory. One of the major results was the identification of the inadequate aggregation methods used by the existing modeling system. He felt his work was handicapped by the lack of homogeneity of loss rates and the instability of aggregated attrition behavior. To thoroughly test his sophisticated shrinkage estimation schemes, and ultimately to apply them, meaningful and well-behaved empirical attrition rates need to be achieved.

Elseramegy [Ref. 6], completed his thesis work on the "CART Program: The Implementation of the Classification and Regression Tree Resubstitution Implementation Application" in December 1985. A goal of his thesis was to apply the CART program to the existing forecasting methods of Marine

Corps officer attrition rates. Ultimately the program proved too difficult for effective use and suffered structural limitations when dealing with cells of potentially widely varying inventories.

Robinson's March 1986 [Ref. 2] thesis, "Limited Translation Shrinkage Estimation of Loss Rates in Marine Corps Manpower Models," was a direct follow-on to Tucker's work. He tested and compared various statistical estimation techniques for the generation of attrition rates. Again, his results revealed the inadequacies of existing officer category aggregations.

Other useful background literature included studies and reports of U.S. Navy issues closely related to this thesis. The work of Siegel [Ref. 7] at NPRDC, describes the seven year attrition rate and forecasting methods used by the Navy. His report describes the Officer Retention Forecast Model (ORFM) and illustrates its capabilities.

A second study done by Bres and Row [Ref. 8] discusses time series-based forecasting techniques used with great success by the Navy in forecasting loss rates within the unrestricted line officer community.

Finally, work by Butterworth and Milch [Ref. 3] presents valuable insight to hierarchical aggregation applications as applied to Navy enlisted ratings.

B. OPERATIONAL AND THEORETICAL BACKGROUND

As this thesis requires a functional knowledge of current and future Marine Corps manpower models the following literature provides necessary operational background.

In the "Functional Description for the Development of the Officer Planning and Utilization System (OPUS)" produced by DSAI [Ref. 4], is provided a written description from the developer to the Marine Corps on the OPUS project. It includes performance requirements of the various models, preliminary design strategies, and user inputs.

The "User's Manual for the Officer Rate Generator," by DSAI [Ref. 9], provides the reader with information necessary for effective use of the officer loss rate generator.

In "System Design for the Marine Corps Officer Rate Projector (MCORP)" by NPRDC [Ref. 5], the MCORP system is discussed in general terms based on operational objectives and design.

The "OPUS--System Specification" by DSAI [Ref. 20] provides a detailed definition of the functions of the Year-Group and Steady-State Promotion models of OPUS.

In "OPUS--System Specification for Optimum Officer Force Model" by DSAI [Ref. 11], an in-depth definition of the functions of the Optimum Force Model and the interfacing

techniques for use with other systems and programs are provided.

The "OPUS--System Specifications for Officer Population Simulation" by DSAI [Ref. 12] defines the functions and details for interfacing the Officer Population Simulator with the planning models of OPUS.

In the "Users Manual for the Officer Planning and Utility System (OPUS)" DSAI [Ref. 13] provides application information for the recently developed Steady-State Promotion and Year-Group models.

A group of textual references address the theoretical concepts as well as the relevant statistical and modeling techniques. These include Bartholomew and Forbes' Statistical Techniques for Manpower Planning [Ref. 14]; Berenson, Levine, and Goldstein's Intermediate Statistical Methods and Applications [Ref. 15]; and Grinold and Marshall's Manpower Planning Models [Ref. 16].

In his classical work on the subject, Johnson [Ref. 17] describes the classical theory and nature of hierarchical clustering as well as illustrative examples of pertinence to this thesis. Further description, discussion and application of cluster analysis techniques and algorithms were provided by Anderberg, Cluster Analysis for Applications [Ref. 18]; Lorr, Cluster Analysis for Social Scientists [Ref. 19]; and Norusis, SPSSX--Advanced Statistical Guide [Ref. 20].

III. DATA BASE

A. GENERAL

The key data base for this analysis is a summary data file designed and compiled by personnel of NPRDC. The summary data file was created from two Marine Corps files: the Headquarters Master File (HMF) and the Quarterly Statistical Transaction File (STATS).

The HMF is the primary source of data for historical officer inventories. September 30 (end of fiscal year) "snapshots," from 1977 to 1986, are used to produce these inventories. The STATS provides input for the generation of historical losses. The two files are merged and sorted to create counts and inventories of all Marine Corps officers of the ten year period [Ref. 5:pp. 2-4--2-22].

The summary data file separates the individual records according to the unique characteristics of MOS, LOS, rank and loss type combinations. The data format is presented in Appendix A.

The summary data file contains a total summary of the actual officer inventory and loss counts of each combination of variable characteristics descriptive of existing officers, by fiscal year. Appendix B provides an example of raw data from the summary data file. The data file is a

direct access file accessible via the Conversational Monitoring System (CMS).

Additionally, the MCORP model, using a flexible multiple-diskette version of the summary data file, allows rapid access to historical inventories and user-weighted loss forecasts through microcomputer application. The MCORP model is capable of generating output in several convenient report formats: Groups by Year, Groups by YCS, and Grade by YCS.

The Defense Manpower Data Center (DMDC), Monterey, California, provided a third source of officer inventory and attrition data. These Defense Department data are essentially similar to those in the summary data file and as a result afford an additional reference resource and an excellent basis for input and output comparisons.

IV. CURRENT SYSTEM AND PRESENTATION OF NEW CONCEPT

A. CURRENT SYSTEM

The historic loss rate calculation is essential to the successful application of the manpower models as emphasized by Barholomew and Forbes [Ref. 14] and Grinold and Marshall [Ref. 16]. Loss rates for OPUS are generated by MCORP from data found in the summary data file. It is the calculation of these loss rates that is hampered by low officer inventories within specific cells, i.e., the small cell problem.

The current approach to answering the small cell problem is termed the "Small Cell Override Methodology" [Ref. 5:pp. 3-10--3-11,H-1]. The goal of the override methodology is to expand the inventories of categories with small populations to avoid over- and under-estimating attrition patterns due to low denominator ratios. As an example, the loss rate resulting from the retirement of one officer during a period, from a population of three (small cell) probably yields a poor base from which to estimate attrition behavior for that group.

Though the data base contains the inventories of ten years, the dynamic nature of officer manpower flows requires that rates reflect current trends as well as long-termed historical attrition. The present procedure is a prototype.

It is acknowledged by user and developer as an interim, ad hoc process based upon perceived officer attrition similarities along traditional classification structures. The specific need for refinement in the small cell aggregation methodology has been demonstrated in the preceding attrition estimation improvement research of Tucker [Ref. 1] and Robinson [Ref. 2].

At present, annual, and even quarterly, loss rate calculations are insufficient to meet the acceptable forecasting tolerances required of the officer manpower planners. Forecasting errors of between 50 and 100 cases occur. The impact such errors have when reconciled with legislated strength authorizations is significant and costly.

With recent emphasis on large scale officer reductions, monthly forecasts are becoming common management requirements. The estimating difficulties encountered with small annual categorical loss inventories are multiplied when faced with monthly estimation demands.

Presently, MCORP offers the user alternative selections of small cell population minimums. Cell inventory specifications are available from one to 50, with a default inventory threshold of 30 cases. This requires that the cell population exceed the specified minimum number of cases. If the cell inventory fails to meet the threshold requirement, the small cell override methodology activates.

A hierarchical series of cellular expansions takes place until the required population is reached. The following paragraphs provide a verbal explanation of the small cell expansion.

1. Test One

Under this test, the single cell is expanded laterally, across YCS, in a stepwise fashion, potentially to include all YCSs[1]. MOS, RANK, and all other variables are unchanged. If this test fails to reach the threshold inventory, then proceed to Test Two.

2. Test Two

With Test Two the single cell is expanded to include all MOSs in the operational MOS Group of the designated MOS. See Table 2 for a description of the traditional MOS groups. YCS, RANK, and all other variables are unchanged. If this test fails to reach the threshold inventory, then proceed to Test Three.

3. Test Three

Using Test Three the single cell is expanded to include all MOSs in its group and YCSs are expanded laterally, in a stepwise fashion, potentially to include all YCSs. RANK and all other variables are unchanged. If this

---

[1]Year 20 is a barrier to YCS expansion from either direction, due to the retirement eligibility. The 20-year YCS is recognized as an obvious boundary of change in loss behavior.

TABLE 2

TRADITIONAL MOS GROUPS

Group Name	MOS
COMBAT	0302
COM/SUPP	0802 1302 1802 1803
COM/SERV	0180 0202 0402 2502 2602 3002 3060 3402 3415 3502 4002 4302 5803
HELO	7562 7563 7564 7565 7566
TACAIR	7501 7508 7509 7511 7522 7523 7543 7545 7556 7557 7576
NFO	7583 7585 7586 7588
AIR/GRD	6002 7204 7208 7210 7820
LAWYER	4402
ALLOTHER	0101 0160 0170 0201 0205 0210 0301 0401 0430 0801 0803 1120 1301 1310 1360 1390 1402 1502 1801 2101 2110 2120 2125 2305 2501 2601 2802 2805 2810 2830 3001 3010 3050 3070 3102 3302 3402 3406 3410 3501 3510 4001 4006 4010 4130 4301 4401 4430 4602 5502 5505 5702 5910 5950 5970 6001 6004 6007 6302 6502 6802 7002 7201 7301 7330 7380 7500 7510 7520 7521 7540 7542 7550 7560 7575 7580 7581 7584 7587 7597 7598 7599 9901 9904 9906 9907 9908 9914 9925

test fails to reach the threshold inventory, then proceed to Test Four.

4. Test Four

Under Test Four the single cell is expanded to include all MOSs. YCS, RANK, and all other variables are

unchanged. If this test fails to reach the threshold inventory, then proceed to Test Five.

5. Test Five

With Test Five the single cell is expanded to include all MOSs, and YCSs are expanded laterally in a stepwise fashion, potentially to include all YCSs. RANK and all other variables are unchanged.

The current small cell aggregation methodology implies several troublesome assumptions. Test 1 expands cells across YCS. The procedure acknowledges the 20-year mark as the single truncation point for significant changes in YCS-based loss behavior. However, in recent years, losses of Marine Corps captains, for example, has taken place over the span of 12 separate YCSs without crossing the 20-year barrier. To assume homogeneous behavior of similarly categorized officers across a broad range of career experience and maturity does not jibe with true attrition rate relationships.

Exploratory clustering of loss rates by YCS for each rank has produced consistent empirical evidence supporting the contention that wide ranges in attrition behavior do occur within classifications based on rank and MOS. Bartholomew and Forbes [Ref. 14:pp. 12-16] discuss the matter of the influence of length of service on attrition rates in more detail.

With the present override, small cells are expanded across MOSs within functionally defined MOS groups in Test 2. This test assumes similar loss behavior among officers in the groups described in Table 1. Do pilots of different fixed-wing aircraft types, group TACAIR, exhibit homogeneous attrition rates? One might expect the job opportunities with civilian airlines to vary between pilots of KC-130 propeller-driven refuelers and pilots of F/A-18 fighter/attack airplanes. Similarly, in the COM/SERV group, highly trained, data systems officers (MOS 4002), with talents readily transferable to the civilian labor market, are aggregated with officers possessing more military specific skills of the intelligence community (MOS 0202). finally, the gross aggregation of the ALLOTHER category combines such diverse groups as basic infantry officers (MOS 0301), disbursing officers (MOS 3402), and student judge advocates (Juris Doctorate in hand, MOS 4401). Though the MOSs in this group tend to be generally rank or YCS specific, loss rates may show excessive heterogeneity in cases where non-MOS categories coincide over diverse occupational specialties.

Test 3 expands cells to include commonly classified officers across all MOSs and YCSs. This aggregation can be characterized as potentially sharing the same assumption weaknesses as the previous stages of the override methodology.

Small cell expansion in Test 4 and Test 5 includes the loss inventories of all MOSs and, in Test 5, all YCSs as well. Though these levels in the hierarchy are infrequently exercised, there appears little theoretical basis to assume that the results of such ranging aggregation might generate particularly homogeneous groups of loss rate behavior.

B. AGGREGATION BY CLUSTER ANALYSIS

This aggregation methodology is proposed in support of an empirical Bayes officer attrition rate estimation scheme under development by Professor R.R. Read. Such schemes utilize the currently popular shrinkage type parameter estimation methods recently researched by Tucker [Ref. 1] and Robinson [Ref. 2].

Statistical methods of this category "shrink" groups of empirical cell rates toward a grand mean. Aggregate rate shrinkage enhances the statistical stability of loss rates, particularly those of small cells. Shrinkage estimation procedures perform best if the designated groups (aggregates) are as homogeneous as possible. It is this final characteristic of internal aggregate homogeneity which led to the application of cluster analysis.

The process of cluster analysis provides an effective tool with which to explore the existing data set for clues about data categorization. In this research the objects of analysis are the specifically classified officer descriptions, i.e., individual cells, in the historical

summary data file. The purpose of clustering is to discover a classification scheme for individual cells which reflects increased homogeneity in attrition rates when compared to traditional groupings.

As described in the previous section, the present aggregation methodology relies on officers in organizationally and functionally defined groups to demonstrate similar attrition behavior. Cluster analysis can lead to the discovery of alternative schemes to the traditional methods of officer categorization. Alternative population partitions which show improved homogeneity of internal historical loss rates can serve as the basis for improved small cell aggregation methods.

Cluster analysis includes many heuristic procedures and statistical applications which can sort data into homogeneous subgroups based on certain measures of similarity. Of application to this study is the hierarchical clustering technique. A brief description of this procedure follows. Greater detail is provided by Johnson [Ref. 17], Anderberg [Ref. 18], and Lorr [Ref. 19].

Hierarchical clustering aggregates objects into sets of clusters according to selected criteria of measured similarity between data elements. A common technique of visual representation of a hierarchical clustering scheme is the dendrogram, see Figure 1.

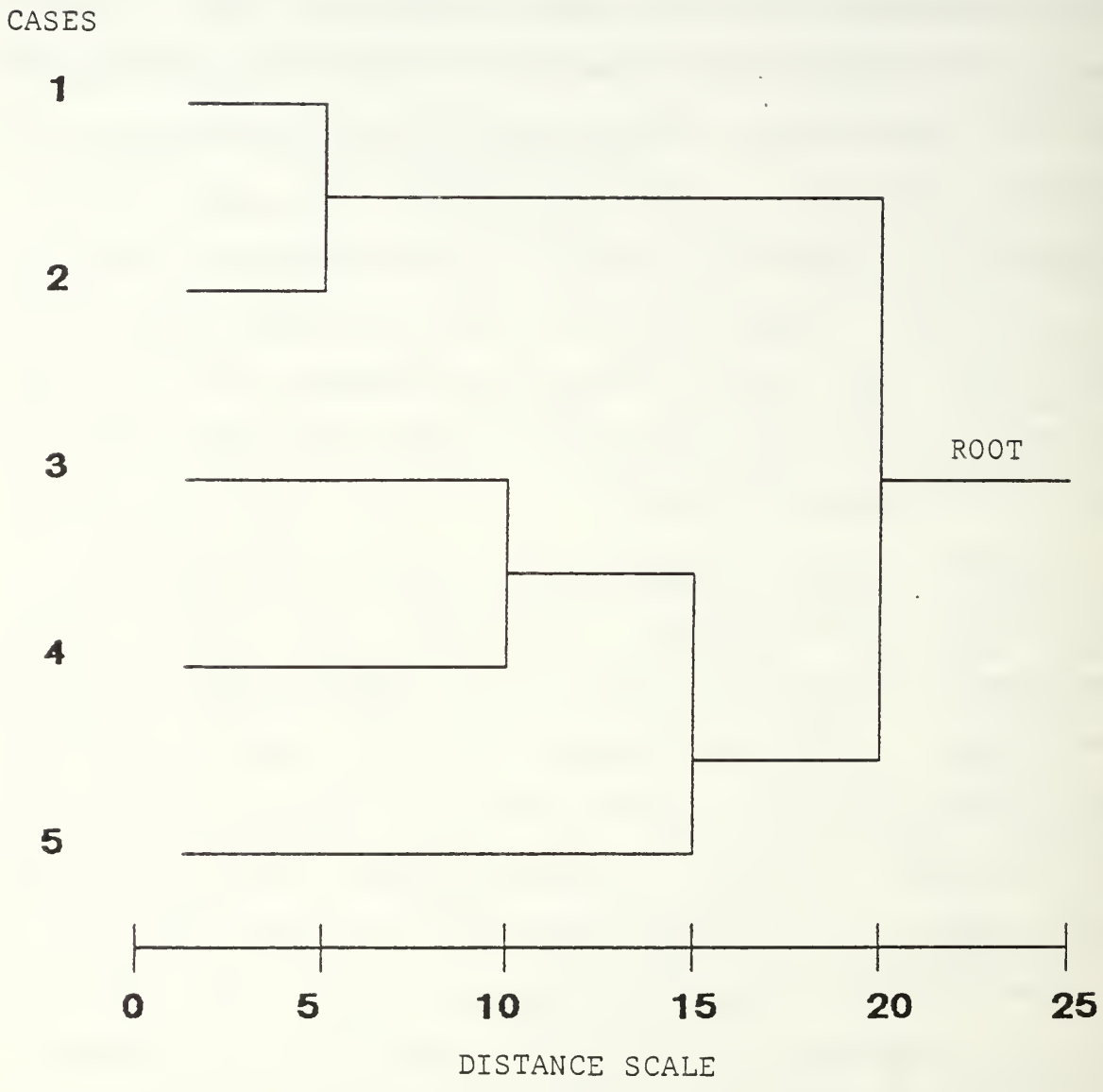


Figure 1. Dendrogram

Cases 1-5 in Figure 1 represent individual objects. The root depicts the aggregation of all objects into one set. By moving from left to right the various entities are sequentially merged into larger and fewer clusters according to the extent of similarity. This is termed the agglomerative method. The distance scale represents the degree of selectivity associated with the formation of the clusters. The smaller the distance, the closer, more similar, are the grouped objects.

In the following sections the major steps in the cluster analysis methodology used in this thesis will be described.

C. DESCRIPTIVE VARIABLES

The variables selected to describe officer attrition are the attrition rates of the loss types as discussed in Chapter I: Retirement, Release, Discharge, Resignation, and Other. This is an inclusive list of both voluntary and involuntary attrition. Summation of the loss inventories equates to the total strength losses.

The data base provides loss counts for each cell in man-quarters, over the ten years of data. Appendix C provides the FORTRAN computer programs used in the creation of various group loss rates. The basic equation used in the computation of annual loss rates is as follows:

$$
\begin{aligned}
l &= \text{annual loss rate} \\
&= 4 \times \text{man-qtr loss counts/year average strength} \\
&= 4L_{ijkm}/(S_{i-1,k} + S_{i,k})^{1/2} \\
&= L_{ijkm}/.125\ (S_{i-1,k} + S_{i,k})
\end{aligned}
$$

where:

l_{ijk} = loss rate of time k, group j, year i

L_{ijkm} = loss inventories in quarter m, type k, group j, year i

S_{ik} = year end inventory of group k in year i.

Within the data file, the annual loss cell inventories have been divided by four for administrative reasons in order to provide planners with quarterly counts. In order to annualize the inventories, the quarterly counts must be multiplied by four, then divided by the total year strength figure. In this case, year strength is an average. As losses take place throughout the year, an average of the beginning and end inventories is the best figure available for total strength.

Two weighted rates were generated for clustering applications. The two rates were computed primarily to facilitate aggregation analysis. Neither rate presumes to reflect the most correct weighting schedule. Such claims are beyond the scope of this research. As a matter of interest the weighted rates were typically very similar, within + or - .005. On the few occasions when the

difference was as much as + or - .01, it was generally an indicator of MOS restructuring due to policy changes or technological advances.

The first rate was an annualized, most-recent-five-year rate recommended for consideration by Professor R.R. Read. The equation follows:

l_{5yr} = five-year average loss rate

= 4 x (sum of man-qtr loss counts, 1982-1986) /sum of yr average strengths, 1982-1986

$$= 4\left(\sum L_{ijkm} / \sum (S_{i-1,k} + S_{i,k})^{1/2}\right.$$

$$= \sum L_{ijkm} / .125\left[\sum (S_{i-1,k} + S_{i,k})\right]$$

In this rate equation the quarterly loss inventories of the five-most-recent-years are summed and annualized (multiplied by four). The result is then divided by the summation of the average total annualized inventories of the same five years.

Such a ratio results in an equal weighting of data from the last five years. The implied presumption of this rate is that an average of recent attrition data provides a better picture of representative strength loss ratios than does any one previous year. Further, that data from years 1978-1981 offer no representative relevance.

A second contrived rate was the weighted loss ratio over the entire data set as recommended by MCORP designer, B. Siegel. The weighting scheme is shown below:

Year	Weight	Ratio
1978	1	.034
1979	1	.034
1980	1	.034
1981	1	.034
1982	1	.034
1983	3	.103
1984	5	.172
1985	7	.241
1986	9	.310

Such a schedule strongly weights the loss data of the most recent years, with earlier years receiving less emphasis. Using this approach, the generally desired preference of utilizing all available data is to an extent realized while giving proportionally greater emphasis on recent activity. The basic equation follows:

l_{10yr} = ten-year weighted loss rate

= 4 x (sum of weighted man-qtr loss counts 1978-1986)/sum of wtd yr average total strengths 1978-1986

$$= 4\left[\sum Wt_i(L_{imjk})\right] / \sum\left[Wt_i(S_{i-1,k} + S_{i,k})^{1/2}\right]$$

$$= \sum wt_i L_{imjk} / .125\left\{\sum\left[Wt_i(S_{i-1,k} + S_{i,k})^{1/2}\right]\right\}$$

where:

wt = weight ratio.

In the ten-year weighted loss rate, the annualized loss inventories are multiplied by a weighting factor prior to summation. (The sum of the weighting factors must equal one.) The weighted sum of loss inventories is then divided by similarly weighted average total strength inventories, summed.

Appendix C again provides a display of the FORTRAN computer programs utilized. To facilitate the application of these rates, the ratios are saved and assembled into a file, in matrix form, by loss type, according to year and specified group. In this situation the specified group provides the clustering elements subject to ultimate aggregation. An illustration is provided below in Table 3 which includes annual rates, the most-recent-five-year average rate (year 98), and the ten-year weighted loss rate (year 99). Appendix D furnishes a complete example of the loss rate matrix.

D. CLUSTERING ELEMENTS

An understanding of the research purpose and initial familiarization with the data set serves as the basis for the development of a clustering strategy. Definitive recipes cannot exist for the selection of clustering

TABLE 3

EXAMPLE LOSS RATES BY YEAR AND MOS GROUP

Year	Group	Retire	Release	Discharge	Resign	Other
78	1	.027	.037	.005	.019	.002
78	2	.029	.049	.006	.021	.002
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
78	12	.064	.012	.003	.005	.001
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
86	1	.021	.053	.003	.016	.001
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
86	12	.003	.005	.004	.006	.000
98	1	.020	.040	.007	.016	.002
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
98	12	.035	.014	.005	.008	.001
99	1	.021	.045	.007	.017	.002
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
99	12	.036	.014	.005	.007	.001

elements which will lead to interesting and relevant classifications. Further, as emphasized by Anderberg [Ref. 18:pp. 182-185], a clustering strategy is generally a

sequential process, responding to increased knowledge about the data and adapting the new information at every stage.

In this study, the clustering elements selected include: YCS, MSO (including various MOS groups), and RANK. Justification for the selection of clustering units follows in the paragraphs below.

Length of service is acknowledged by Bartholomew and Forbes [Ref. 14:p. 14] and others, as a primary, if not dominant, factor affecting the propensity of an individual to leave an organization. In general, the propensity to leave decreases with increased length of service, salary, and status.

In this research, YCS is used as a surrogate for length of service. This substitution appears appropriate as, in the large majority of cases, YCS equals length of service. In the relatively infrequent situations where unrestricted officers have significant amounts of enlisted service, YCS is less than actual length of service. In these cases, however, YCS is still a major determinant for promotion, authority, and responsibility.

In the loss rate matrix formation, YCS (one year to 31 years) becomes a specified row identity. Appendix C provides the computer program utilized and Appendix D offers the loss rate matrix file. YCS is initially clustered over the entire data set for a broad perspective of data loss

rate behavior. Subsequently, YCS is clustered with respect to more homogeneous MOS groups for comparison and analysis.

The 140+ officer MOSs in the Marine Corps represent a diverse collection of fields and duty descriptions. MOSs vary in the amount and expense of initial and follow-on training required to fulfill occupational requirements. As a result, varying degrees of transferability of skills to the civilian labor market can be identified with MOS categorization. The training required of a lawyer (MOS 4402) or basic jet-fighter pilot (MOS 7520) is far more expensive in time and money than initial training for an officer in the intelligence specialty (MOS 0202). Further, the value of equally transferable skills can also vary. Both a multi-engine KC-130 aircraft pilot (MOS 7557) and a military police officer (5803) might share easily transferable skills but the corresponding civilian salaries for similarly successful former officers may be quite different.

Some specialties exhibit more typically arduous duties, such as infantry (MSO 0302) or combat engineer (MOS 1302). Such differences may be reflected in the collective attrition behavior. Still other specialties may be identified as quite unique in a variety of obvious and less than obvious characteristics of duty, population, or environment which cause them to respond with significantly different group loss rates.

Due to the mentioned theoretically-based variances and differences, less well-understood or accepted, even the casual observer would expect divergent attrition behavior across the various MOSs and MOS groups. MOS appears to be a logical and appropriate clustering variable which intuitively should yield interesting categorizations.

Three variations of MOS groupings already exist in functional hierarchy. The lowest level is the four-digit MOS. The next degree is the occupational field group. These groups consist of all MOSs sharing similar first and second digits. Occupational field (OCCFLD) 34, Auditing, Finance, and Accounting, consists of MOS 3401, MOS 3402, and MOS 3415. Finally, the MOS groups described in Table 2 are the largest of the three groupings. These three categories provide the initial clustering elements for analysis.

Similarly to YCS, MOS or MOS groups become the row identities in the loss rate matrix formation. Appendix C provides the computer program utilized in this project and Appendix D offers the subsequent loss matrices.

RANK was the third designated clustering element. This characteristic is strongly associated with YCS but does offer a measure of officer performance. The utility of RANK as a performance measure is enhanced by the inclusion of failed-select status as a categorization. For the complete categorization of RANK, see Appendix A.

Since the scope of this thesis does not cover warrant officers or LDOs, these ranks are eliminated from the RANK clustering. Appendix C includes the programming of the RANK variable.

For the manpower manager, the selected clustering elements represent the most interesting descriptive aspects of the officer populations with regard to attrition behavior. RANK, YCS, and MOS are the major elements of management concern and are the natural cases to be used in the definition of new attrition rate aggregates.

E. SIMILARITY MATRIX

The hierarchical clustering method requires that every pair-wise combination of clustering variables be defined by a measure of similarity. Similarity is measured by the proximity or distance between entities. The process of similarity computation leads to the creation of a lower triangle similarity matrix. Figure 2 shows the similarity matrix.

There exist numerous distance measures available for use in the creation of the similarity matrix. Lorr [Ref. 17:pp. 32-34] and Anderberg [Ref. 18:pp. 98-110] discuss various distance functions referred to as metrics. The Chebychev distance metric was selected as the measure for use in this research and is represented as follows:

$$D(x,y) = \text{MAX} \, |X_{ij} - Y_{ik}|$$

where:

X_{ij} = loss rate of the jth cell of the ith variable

Y_{ik} = loss rate of the kth cell of the ith variable.

$$
S = \begin{matrix}
 & & & & \\
S_{21} & & & & \\
S_{31} & S_{32} & & & \\
S_{41} & S_{42} & S_{43} & & \\
\vdots & \vdots & \vdots & \ddots & \\
S_{n1} & S_{n2} & S_{n3} & \cdots & S_{n(n-1)}
\end{matrix}
$$

Source: M.R. Anderberg, Cluster Analysis for Applications (New York: Academic Press, 1973): 133, Figure 6.2.

Figure 2. Lower Triangle Similarity Matrix

The Chebychev metric measures the distance between entities as the maximum absolute difference in value for any one variable. When officer attrition behavior is characterized by the previously mentioned rates of loss, typically it is one, or perhaps two, of the rates that are of interest at a particular career moment. It is these

singular rates that practically define the unique nature of the individual cell. The outlying, or distinguishing, loss rate is the primary ratio of interest that is best isolated by using the Chebychev metric. For instance, the most-recent-five-year loss type attrition rates for an infantry captain (MOS 0302) are:

Loss type:	Retire	Release	Discharge	Resign	Other
Loss rate:	.002	.012	.002	.030	.001

Compare these to the rates of an aviator captain, who flies F-4 fighter aircraft:

Loss type:	Retire	Release	Discharge	Resign	Other
Loss rate:	.000	.053	.002	.032	.003

In this situation the Release type loss rate is the aspect of attrition that distinguishes the otherwise similar loss behavior difference between these two categories of officers. The Chebychev metric bases the calculation of similarity on the maximum difference of loss rate types. Thus the nature of the data suggests the Chebychev distance metric.

Alternatives to the Chebychev metric often are based on the sums of differences between variables and would obscure the most dramatic aspects of cell differences. Further, the fact that the loss variables are of a binomial distribution leading to unequal variances causes all measures based on

Euclidean distance to be inappropriate, e.g., squared Euclidean distances, Manhattan distances, etc.

A sample calculation using the Chebychev distance metric is given below:

78	1	.027	.037	.005	.019	.002
79	2	.029	.049	.006	.021	.002

$$D(1,2) = MAX_i|X_{ij} - Y_{ik}|$$

$$= MAX\ |.027-.029|,\ |.037-.049|,\ |.005-.006|,\ |.019-.021|,\ |.002-.002|$$

$$= MAX\ \ .002,\ .012,\ .001,\ .002,\ .001$$

$$= .012$$

The hierarchical clustering technique is executed over the similarity matrix constructed of resultant distance measures. The SPSSx program allows for the specification of the Chebychev distance metric by subcommand in the procedure CLUSTER as described by Norusis [Ref. 20:pp. 184-185].

F. CLUSTERING CRITERION

Once the similarity matrix is defined, the choice of clustering criterion must be addressed. Clustering criterion describes how the most similar clusters are to be selected. This is the computational burden of the hierarchical clustering technique.

Both Lorr [Ref. 17] and Anderberg [Ref. 18:pp. 134-145] offer a variety of clustering criterion options. Every clustering method is nominally unique and apart from every other method. However, many of the methods tend to yield substantially similar results.

The method selected for this research is known as the average between group method. This technique evaluates the potential merger of all clusters in terms of the average similarity of the links between the cluster pairs.

Initially, several alternative schemes were rejected as inappropriate due to association with various squared Euclidean distance metrics. Further, the simplest linkage methods tend to base clustering decisions on the minimum or maximum distance cluster membership, e.g., the single linkage and the complete linkage methods. To avoid such dependency on extreme values for the definition of clusters, a method using the average of all links of cluster pairs was considered most useful and correct. Two such methods are the average linkage between groups and the average linkage within groups:

Average linkage between groups

$$\frac{SUM_i + SUM_j + S_{ij}}{(N_i + N_j)(N_i + N_j - 1)/2}$$

Sum i = sum of an pairwise similarities among entities within cluster i

N_i = the number of entities in cluster i

Average linkage within groups $\frac{S_{ij}}{N_i N_j}$

No theoretical considerations or technical explanations offer sufficient reason to select one method over the other. A test was therefore constructed to compare the clustering solutions using the two candidate methods.

Twelve sets of seven or nine pairs of numbers from 0 to .50 were generated to simulate loss rates. The sets were clustered using the SPSSX CLUSTER procedure and the results were plotted for comparison.

As anticipated, the majority of the comparisons showed little if any difference in aggregation hierarchy. However, a few of the sets did show distinct differences and demonstrated important clustering trends. The average linkage within groups tended to cluster one or two distinct groups initially and quickly expand the existing clusters into higher levels of aggregation. The average linkage between groups tended to create more clusters initially and pool clusters into higher levels of aggregation later in the sequence.

More clusters at the lowest level of an agglomerative hierarchy provide greater insight into data set relationships characterized by inherently small ratio differences. The tendency to establish more clusters initially was consistent with the needs of this project. Therefore, the clustering exhibited in the average linkage

between groups was preferred and the between groups method was selected as the criterion for clustering. The test is documented in Appendix E.

The SPSSx program allows for the average linkage between groups method to be specified by subcommand in the procedure CLUSTER as offered by Norusis [Ref. 20:pp. 184-185].

G. DENDROGRAM

A final aspect of the hierarchical clustering analysis concerns the clustering result. As indicated earlier the dendrogram offers a convenient display of the clustering sequence and composition. It is desirable in this work to also measure the relative population sizes of the clusters represented in the aggregation.

A separate program was created to calculate the cumulative population of the associated officers with each stage of the aggregation. Using SPSSx the calculation of cluster membership at specified stages of aggregation can be accomplished. Appendix F provides the program utilized.

V. APPLICATION OF CLUSTER ANALYSIS

A. GENERAL

Using the hierarchical cluster analysis methodologies and techniques described in the preceding chapter, loss rate matrices and dendrograms were computed and drawn for a variety of clustering strategies. Introductory loss rate clustering was conducted on the MOS groups from Table 2, as well as the 47 OCCFLDs. The attrition rates of the MOS groups appeared to cluster as expected with aviation-type groups together and ground-type groups together, etc. However, when the loss rates of the OCCFLDs were clustered, over the entire population, unexpected relationships developed and many perceived similarities were found to be without statistical support.

Length-of-service as discussed previously may be viewed as the driving force behind attrition behavior. YCS, a length-of-service surrogate, was clustered over the entire population. The results of YCS aggregation demonstrated significant and consistent attrition is associated with various lengths of service. See Figure 3 for illustration of this point.

On inspection, the general, all-service YCS aggregation in Figure 3 was found credible. Year four, for instance, stands out as a distinct YCS quite in terms of attrition

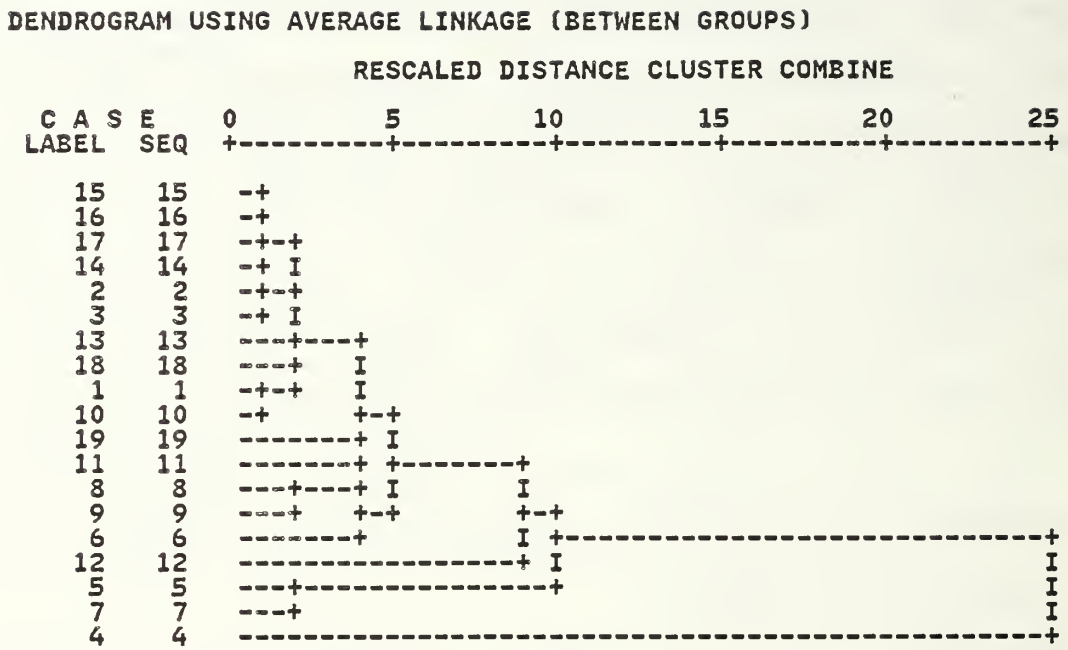


Figure 3. Entire Population YCS Aggregated Dendrogram

behavior. The loss rates of Marine officers in their fourth YCS do not cluster with other YCSs until the final stages. This follows on the basis that the fourth YCS is the time when the majority of initial service obligations are met and a relatively large number of officers elect to leave the Marine Corps. The appearance of the YCS aggregation dendrogram follows well the factors of service obligation, selection for promotion, and retirement opportunities.

Exploratory clustering of RANK in various combinations across a variety of MOSs substantiated that attrition behavior is strongly associated with specific rank. Each level exhibited its own unique characteristics. Loss rates of captains were generally similar, attrition rates of failed-select majors were basically the same, etc.

Promotion to higher rank is largely a function of YCS and in practical terms demotion does not exist in the Service. Further, accelerated promotion seldom occurs and the advancement of officers through the rank of lieutenant colonel is fairly predictable. For these reasons RANK was not selected as an element for further cluster analysis. The situation of failed-selectees can be adequately addressed by the designation of the failed-select categories in cell definition, see Appendix A.

With examination and comparison, the above clustering schemes led to the discovery of various relationships and the development of still more clustering approaches.

Inevitably specific loss rate case outliers were encountered which did not neatly fit into specified groups. In the interest of time and expense, those that nearly qualified were most often subjectively included into existing groups. Outliers with great dissimilarities were individually identified, investigated and as necessary, isolated.

The proposed replacement for the current small cell override methodology is presented in Table 4. A discussion of the development of this solution is embodied in the remainder of this chapter.

B. SMALL CELL DEFINITION

Prior to addressing the aggregation specifics, the small cell population threshold warrants attention. The small cell population threshold is the factor that determines the extent of aggregation which will occur when a small cell is encountered in the course of a problem involving MCORP.

The small cell population should remain a flexible aspect of the MCORP model. The ability of the user to specify a minimum small cell population is a desirable feature of this process. Such control can be used to influence the conservativeness of small cell loss rate generation.

Selection of a low inventory threshold results in rates reflective of relatively few observations in a narrow range of parameters. The potential for accurate loss estimation from these values exists but the risk of gross error in the

TABLE 4

PROPOSED SMALL CELL OVERRIDE METHODOLOGY

Stage One

Expand selected cells incrementally, by YCS: (YCS+1), (YCS-1), (YCS+2), etc., until the boundaries below are reached.[1] Stop when the population threshold is met or exceeded:

MOS Category	Bounded YCS Groups[2]			
Fixed-Wing Pilots	(1-6, 8-19)	(7)	(20-25)	(26)
Rotary-Wing Pilots	(1-5, 8-19)	(6,7)	(20-25)	(26)
Naval Flight Officers	(1-5, 8-19)	(6,7)	(20-25)	(26)
Lawyers	(1-6, 8-19)	(7)	(20-25)	(26)
All Else	(1-3, 6-19)	(4,5)	(20-25)	(26)

If inventory is below the threshold retain the accumulated inventory and continue to Stage Two.

Stage Two

Expand the cell resulting from Stage One expansion to include the specified YCS in the Small MOS Groups defined below[3]. If inventory is yet below the threshold, expand the YCS incrementally as in Stage One. Stop when the threshold is met or exceeded.

---

[1]See the YCS expansion example in Section C of this chapter.

[2]YCSs beyond 26 are not addressed in this work.

[3]See Section D of this chapter for more details of cell expansion.

TABLE 4 (CONTINUED)

Small MOS Group		MOSs							
FWP	=	7501	7511	7522	7542	7543	7545	7576	
BCP	=	7500	7510	7520	7530	7540	7550	7560	7575
P/RP+	=	7551	7552	7555	7556	7557	7559	7562	7565
		7566	7587						
NFO+	=	7581	7583	7584	7585	7586	7588	7508	7509
		7563	0202	3415					
STDA	=	7580	7597	7598	7599				
CMBT	=	0302	0802	1302	1802	0180	2602	7204	7210
		7220	7320	7564					
SUP1	=	0402	1803	2502	3002	3060	4401	4402	5803
		7208							
SUP2	=	4002	4302	6002	6102				
STDG	=	0101	0201	0301	0401	0801	1301	1801	2501
		2601	3001	3401	3501	4001	4301	5801	6001
		6101	7201	7301	9901				
F-18	=	7521	7523						
MTO+$	=	3402	3502						

If inventory is below the threshold retain the accumulated inventory and continue to Stage Three.

Stage Three

Expand the cell resulting from Stage Two expansion to include the specified YCS in the Large MOS Groups defined below. If the inventory is yet below the threshold, expand the cell incrementally by YCS as in Stage One. Stop when the threshold is met or exceeded.

Table 4 (CONTINUED)

Large MOS Group		Small MOS Group
PLT+	=	FWP, P/RP+, SUP2, NFO+
STD	=	STDA, STDG, BCP
F-18	=	Expand no further, accept the rates with due regard for low inventory. Rates are flagged to indicate inventory of less than threshold.
GRD	=	CMBT, SUP1
MTO+$	=	Expand no further, accept the rates with due regard for low inventory. Rates are flagged to indicate inventory of less than threshold.

If the inventory is below threshold retain the accumulated inventory and continue to Stage Four.

Stage Four

Expand the cell resulting from Stage Three expansion to include the specified YCS in the Major MOS Groups defined below. If the inventory is yet below the threshold, expand the cell incrementally by YCS as in Stage One. Stop when the threshold is met or exceeded.

Major MOS Group		Large MOS Group
AVN+	=	STD, NFO+, PLT+
GRD	=	GRD

If the inventory is still below the threshold, expand no further. Accept the rates with due regard for the low inventory. The rates are flagged to indicate inventories of less than the threshold.

estimation of true attrition behavior increases dramatically when reliance is made on the actions of only a few individuals.

The selection of a high threshold inventory value yields aggregations of small cells at a higher level, diluting the unique historical loss and strength inventories of the cell with populations of other cells of theoretically somewhat less-similar behavior.

The current MCORP design allows the user to select a minimum small cell inventory threshold of from one to 50. It seems reasonable that a higher threshold may at times be of interest to the manpower planner. Small cell definitions of 100 or even 250 may be useful in the course of typical manpower problem analysis and investigation.

C. STAGE ONE--YCS EXPANSION

Stage One of the proposed small cell aggregation solution relies on the dominance of length-of-service characteristics on officer attrition behavior. Small cells are expanded incrementally by YCSs within specified boundaries while all other cell characteristics remain unchanged.

Each MOS is identified with a specific set of YCS boundaries. The example below provides an illustration of the YCS expansion of a cell categorizing A-6 (fixed-wing aircraft) pilots (MOS 7518). This MOS would be identified with the YCS boundary for fixed-wing pilots.

YCS	Incremental Expansion			
	Range of 1st Expansion	Range of 2nd Expansion	...	Range of Max Expansion
4	4,5;	3,4,5;		1,2,3,...6,8,9,...19
7	no expansion	no expansion	...	no expansion
20	20,21;	20,21,22;	...	20,21,...25,27,28,29
26	no expansion	no expansion	...	no expansion

Figure 4 provides the supporting dendrogram for fixed-wing jet pilots clustering YCSs 1-26.

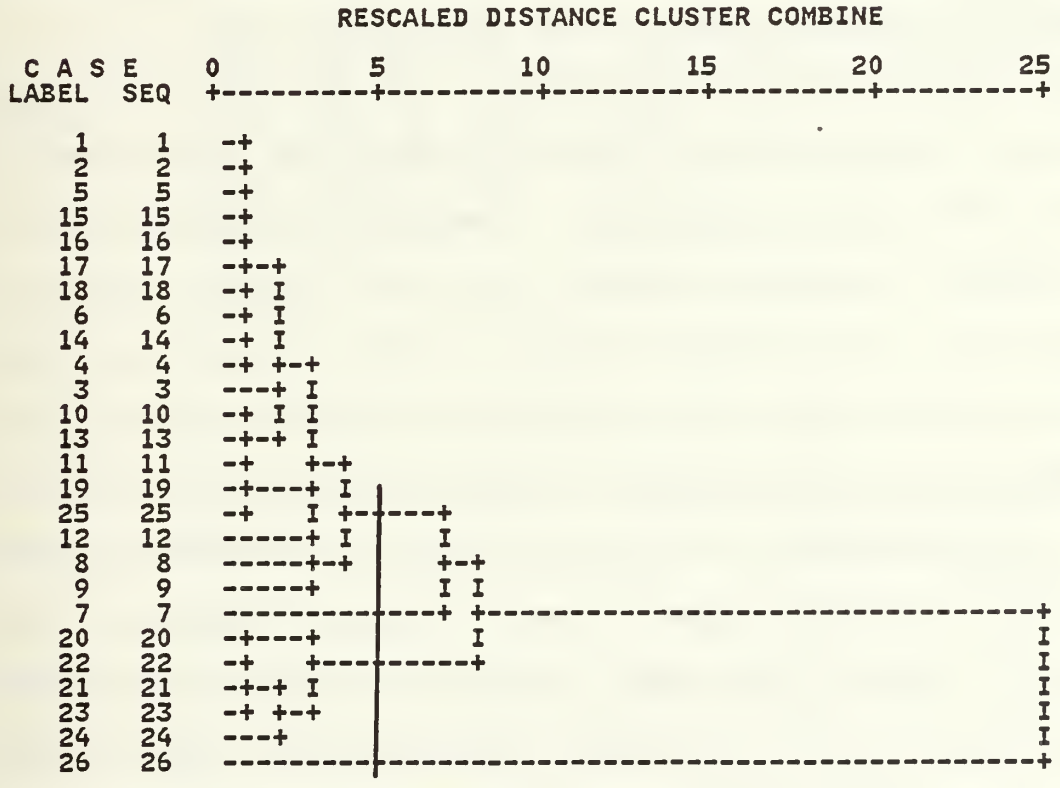


Figure 4. Dendrogram Fixed-Wing Plots--YC

The selection of the YCS boundaries requires a degree of subjectivity. The solid, vertical line drawn through the dendrogram in Figure 4 defines four YCS groups. This line can be shifted left or right to define more or fewer clusters. The farther left the shift the more homogeneous are the separate groups.

Not all dendrograms are as easily applied as Figure 4. Often subjective decisions must be made when the results of dendrograms describing various MOSs and MOS groups are compared. Ultimately each MOS was identified with one of the five common YCS cluster schemes associated with the MOS categories listed in Stage One on Table 4.

The distance scale across the top of the dendrogram provides a rescaled range which reflects the ratios of the computed distance coefficients. (Recall the discussion of the Chebychev distance metric and the linkage method in Chapter IV.) Of greater utility in this project than the distance coefficients was a cumulative inventory of officers associated with the clusters as they aggregate. For this reason, a supplemental table was created to show the summed inventories of each cluster at every stage of the clustering hierarchy. Such an inventory provides valuable insight into the relative sizes of various clusters which aids in the required subjective decisions concerning the significance of various aggregation approaches. The selection and comparison of YCS boundaries and all aspects of cluster

analysis in this thesis were dramatically enhanced by the development of the strength tables.

Table 5 is an example of the inventory strength table of the final stages of aggregation which supplements the example in Figure 5. The computer program is provided in Appendix F.

TABLE 5

CLUSTER STRENGTH TABLE

FIXED-WING PILOTS--YCS

number of clusters ...	12	11	10	9	8	7	6	5	4	3	2
cluster inventories	542	749	749	877	877	944	1006	1173	1284	1284	1402
	128	128	128	111	111	111	111	111	113	118	4
	111	111	111	87	167	167	167	113	5	4	
	87	87	87	80	62	62	113	5	4		
	80	80	80	62	67	113	5	4			
	207	62	62	67	113	5	4				
	62	67	67	113	5	4					
	67	74	113	5	4						
	74	39	5	4							
	39	5	4								
	5	4									
	4										

The designation of rank in the classification process will typically limit the expansion of YCSs to ranges less than the theoretical limits of the Bounded YCS Groups defined in Table 4. If a cell containing lieutenant colonels of a particular MOS, in their 17th YCS, was defined as a small cell, the maximum YCS expansion may only include the 16th, 18th, and 19th YCS since all other YCS cells may be empty. This could happen since few officers are

currently promoted to lieutenant colonel prior to the 16th YCS.

A schedule for YCS expansion beyond 26 YCS is not included in this study. Attrition rates for officers beyond the 26th YCS are not a major planning concern. Though trends do exist in these upper ranges, they are not nearly as well behaved as are the aggregates from one to 26. In general, however, the 27 to 29 YCS rates aggregate early, indicating similarity, as do the 30 and above rates.

Once the YCS expansion meets or exceeds the designated inventory threshold, the attrition rates are calculated according to a user-defined weighting scale. This weighting scale allows for variable emphasis to be placed on the attrition activity of a specific year, or years, of observations in the data base. The loss rates are then available for immediate use or for forecasting applications. If the designated threshold inventory is not attained with the maximum expansion of YCSs, the same cell is further expanded according to the methods described in Stage Two.

D. STAGE TWO--SMALL MOS GROUPS

In Stage Two, small cells are expanded by MOS to include the Small MOS Groups defined in Table 4. Initially, the expanding small cell inventories will reflect only the YCS described in the original officer description (small cell classification). That is, the small cell inventory, following Stage One, is increased first by the inventories

of similar cases with the specific YCS in each of the MOSs of the Small MOS Group.

If the cell population remains below the designated threshold, the cell is further expanded incrementally, by single YCS, over the entire Small MOS Group according to the YCS expansion sequence described in Stage One. This process allows for a gradual increase of inventories and implies that those populations in similar MOSs are most similar nearer the original small cell YCS. If YCS expansion within the Small MOS Group reaches the limits of the Bounded YCS Group without meeting the small cell threshold, Stage Three expansion is conducted as described in Table 4.

Clustering of MOS groups is a logical, next level, association. Historically, officers of different specialties tend to exhibit variations in group attrition behavior. Obvious reasons for such group differences include variable service obligations associated with MOS training, transferability of acquired military skills to the civilian labor market, career potentials identified with specific specialty groups, etc. Less obvious reasons for differences in group loss behavior may include MOS assignment practices, differences in the extent of family separations, and other variable characteristics which are MOS or MOS group specific.

As mentioned previously, the current small cell expansion design aggregates MOSs by similar functional

characteristics of the specific occupation: fixed-wing aviators, combat, combat service, combat service support, etc. The proposed methodology expands small cells by including the populations of similar cells of other MOSs which exhibit similar attrition rate behavior into the total inventory. In order to discover homogeneity in loss rates among all the MOSs, exploratory aggregations were performed on MOSs, OCCFLDs, and the existing functional MOS groups.

MOS groups as currently structured were found lacking in homogeneity of attrition rates, particularly in the non-aviation specialties. MOS groups often contained broad loss rate variations reflecting conspicuous differences in attrition rate behavior. When the current MOS groups were clustered and studied, various subsets were discovered which later formed the nucleus of new Small MOS Groups. Even within the more narrowly defined aviation specialties, further categorization appeared appropriate.

When the 47 OCCFLD groups were clustered, several counter-intuitive aggregations were formed. On investigation it was discovered that the inclusion of basic officers in the OCCFLD groups often distorted the collective loss rates of the OCCFLD groups. Basic officers are less-than-fully-trained officers in a specialty normally associated with their ultimate MOS. They are designated by the third and fourth MOS digits "01." See Appendix A, Primary MOS. By including the basic officers in the loss

rate calculations of the specific OCCFLDs the more meaningful and important attrition similarities of the fully trained officers were masked. As a result the groups defined by OCCFLD definition were found fully lacking in desired homogeneous loss rate behavior and were generally disregarded as indicators of similar attrition behavior.

Following the cluster analysis of the functional MOS groups and OCCFLDs the loss rates of all individual MOSs were aggregated. In this way, clusters of historically homogeneous loss rates were developed without regard to previously accepted aggregation groups.

The results were generally intuitively agreeable. Strong rate similarities existed among various aviation elements presumably attributable to some extent to common service obligations resulting from time spent in flight training as well as the lucrative civilian market-value of a trained aviator. Equally pronounced relationships were exhibited by the non-aviation specialties.

The initial service obligations are generally similar among the non-aviation MOSs. Some vaguely similar clusters did develop around the previously described functional groups of combat support (COMM/SUPP) and combat service support (COM/SER). However, many of the loss rate differences appear to coincide with the occupational transferability to the civilian labor market. Note the example of the diversity of MOSs clustered into the Small

MOS Group SUP1. These include communications officers (MOS 2502), assault amphibian vehicle officers (MOS 1803), air support control officers (MOS 7208), and lawyers (MOS 4402). The functional differences of these specialties is large, however, the relative ease of transferability of skills to the civilian labor market is quite similar.

From both the aviation and non-aviation areas, the MOSs of basic officers clustered nicely into distinct loss rate behavior groups. Intuitively, one might expect officers yet to attain fully trained status to show more similarity in attrition behavior as a group than with officers of acknowledged skill level and accomplishment.

Similarly, within the aviation community, strong aggregation was found among basic pilots. Basic pilots are aviators who have completed flight school but have yet to attain a proficiency for a particular aircraft type.

The inevitable outliers were encountered in the development of this and subsequent aggregating schemes. It was in the handling of outliers that a degree of subjectivity, organizational knowledge, and intuition were frequently exercised. Some MOSs had very small or zero loss inventories. Rather than aggregating these cells together, as would be done with the clustering procedure, they were grouped with similar specialties based on the researcher's knowledge of function, initial and subsequent training requirements, and length of initial service obligation.

Such generalizations are not expected to lead to methodology execution difficulties.

One special case should be noted here. The inclusion of CH-53 helicopter pilots (MOS 7564) with CMBT seems contrary to the general scheme. The inventories available for this MOS are, however, of sufficient magnitude to support the verity of such deviant behavior. This type of situation needs special recognition when the officer planner is making germane analytical decisions.

The case of the motor transport officers (MOS 3502) and disbursing officers (MOS 3402), is however handled differently. This is a relatively large MOS group which includes nearly three percent of the officer population. The attrition behavior of this group is dramatically different from all others. As a result the MOSs typically cluster with other groups during only the last stages of aggregation. Investigation of the loss rates revealed consistently high Release loss rates probably attributable to past assignment policies. As long as such differences exist, the MOS is best treated in isolation, that is, given a unique Small MOS Group. Such special handling demonstrates the need for system flexibility to manage the changing environment in order to meet the routine needs of the manpower planner.

The other example is the situation of the F-18 pilots. The loss rates of these pilots (MOS 7523) aggregate with

others in the very late stages of clustering due to the high rate of Resignations for this group in 1985 and 1986. Without understanding the causes of such trends, it appears better to isolate the MOS as a unique group for the present time and rely on the judgment of the officer plans analyst to properly apply the atypical attrition behavior of this group.

E. STAGE THREE--LARGE MOS GROUPS

In Stage Three, small cells are expanded to include similarly characterized inventories in the Large MOS Groups defined in Table 4. Initially the inventories will reflect only the YCS(s) described in the original classification. If the cell population remains below the designated threshold, the cell is enlarged according to the YCS expansion sequence appropriate to the specified MOS(s) as in Stage One. If, in the end, the small cell still exists and further expansion is possible, Stage Four expansion is conducted as described in Table 4.

The Large MOS Group classification represents an increasingly higher aggregation of loss rates within an MOS hierarchy. Figure 5 illustrates the aggregation of the Small MOS Groups. The solid vertical line drawn through the dendrogram of Figure 5 defines the Large MOS Groups. ELSE includes MOSs of unrestricted officers and may be disregarded.

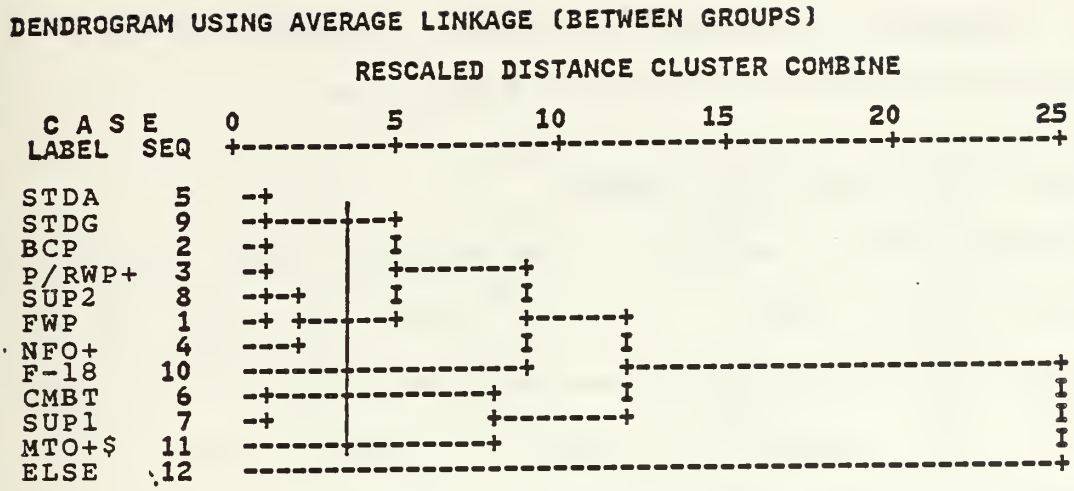


Figure 5. Dendrogram--Small MOS Groups

The first cluster aggregates basic officers of aviation and ground communities with basic pilots. In this case, the initial service obligations seem less important to loss rate behavior than are the similarities of rank, age, career motivation, etc. As few in these groups reach the initial career decision point while in a basic status, this aggregation appears quite rational.

Loss rates of pilots and NFOs aggregate at this stage as might be expected but they are also joined by the small category of SUP2 which contains data system officers (MOS 4002), public affairs officers (MOS 4302), and aircraft maintenance officers (MOS 6002 and 6102). Peculiarities of initial military training in these highly specialized fields may explain this association.

The CMBT and SUP1 groups cluster early and represent well the bulk of the ground MOSs of fully trained officers (recall the exception of CH-53 pilots). The MTO+$ and F-18 outlier groups remain as isolated aggregates.

F. STAGE FOUR--MAJOR MOS GROUPS

In Stage Four, a portion of the enduring small cells are expanded a final time. At this stage PLT+ and STD are aggregated, while the ground occupations (GRD) join F-18 and MTO+$ at the expansion limit. Again, the initial aggregation of small cell inventories will reflect only the YCS(s) described in the original classification. If the cell population remains below the designated threshold, the cell is enlarged according to the YCS expansion sequence appropriate to the MOS(s) as described in Stage One. Due to the remaining diversity in the final group rates it appears to be more appropriate to accept the rates of the low inventories below threshold, rather than distort them further by continued aggregation. See Table 6 for rate differences.

In Figure 6 is provided illustration of the Large MOS Group aggregation. The solid line indicates the definition of the Major MOS Groups.

TABLE 6

LARGE MOS GROUP LOSS RATES

loss type: Group	Return	Release	Discharge	Resign	Other
PLT+	.010	.008	.006	.014	.002
STD	.023	.034	.009	.027	.003
F-18	.013	.027	.001	.065	.004
GRD	.022	.063	.010	.024	.001
MTO+$	.010	.107	.011	.020	.001

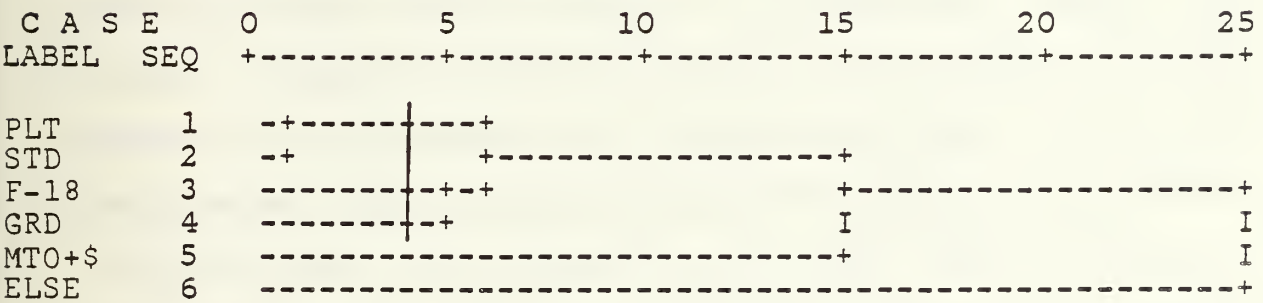


Figure 6. Dendrogram--Large MOS Groups

VI. CONCLUSION

A. SUMMARY AND CONCLUSIONS

The purpose of this study was to aggregate low inventory, officer categories (small cells) into sets of homogeneous historical loss rate behavior. The selection of the hierarchical clustering method in lieu of current functional and organizational structures for aggregate formation supports an empirical Bayes rate estimation scheme currently under development by Professor R.R. Read. Both the aggregation method and the rate estimation scheme are parts of a large effort concerning model development and system integration under the broad title of OPUS.

This thesis presents an algorithm for the aggregation of small cell populations of Marine Corps, active duty, officers in the ranks of second lieutenant through colonel. This project has demonstrated the adaptability of cluster analysis to officer attrition rate aggregation and provides an initial shell for more refined model applications. Further, statistical stability and loss rate homogeneity have been introduced to allow more successful application of powerful shrinkage type parameter estimation methods.

Several points of interest can be identified. First, the present aggregation of small cell populations over YCSs fails to associate periods of initial service obligation

completion to loss rate variability. The years when most officers of a specified group fulfill their initial service contracts are characterized by loss rates greatly different from the rates of adjacent years not coinciding with initial career decisions. Attrition behavior homogeneity is enhanced when these periods are isolated. Similar criticism holds for other key career decision points later in the range of YCSs.

Secondly, existing MOS aggregations of officer loss rates vary, often radically, from groups developed using cluster analysis techniques. This is particularly apparent in the non-aviation MOSs. In many cases the transferability of military skills appears to be of greater influence on attrition behavior than the traditional organizational or functional categories. Also, the extent of initial training and service exposure manifests as an important influence in loss rate behavior. Cluster analysis allows for the segregation of different rate behavior groups by MOS without the burdens of rationalizing why such differences exist.

Finally, there appears to be a logical barrier to the limits of officer small cell loss rate aggregation. A level exists, albeit subjectively, where continued aggregation seems to be counter-intuitive. Some groups of attrition rates are so different from others that to join them would either improperly mask the identity of the small group or severely distort the unique characteristics of two similarly

sized sets. In cases as these, the small cell rate utilization is best left to the judgment of the manpower planner and analyst.

B. RECOMMENDATIONS

It is recommended that the cluster analysis aggregation scheme herein be considered as a prototype for the development of an algorithm to be programmed for small cell aggregation within MCORP. It is anticipated that the application of this scheme to the small cells of Marine Corps warrant officers and LDOs can be accommodated with relative ease. Adoption of this method will allow further research and ultimate application of the desired shrinkage type parameter estimation methods to OPUS.

The current MCORP design allows the user to select a minimum small cell inventory threshold from one to 50. Loss inventories associated with those cells are typically far less: generally about one-tenth of the population. The significance of even one or two losses is therefore profound on the resultant rate. In the interest of conservative applications, it is recommended that the maximum threshold for small cell definition be raised to 100 or even 250 observations. It is anticipated that such a ceiling would be useful in the course of routine manpower analysis and investigation.

Finally, as this is a dynamic system, a commitment to periodic reevaluation and maintenance must be accepted. An

annual methodology update, or special updating procedures as required by significant policy changes, could be easily programmed and would ensure the maintenance of current attrition rate relationships.

APPENDIX A

DATA FORMAT

This appendix contains the summary data file format. The source of this file format is B. Siegel, the program designer, Navy Personnel Research and Development Center, San Diego, California.

FILE FORMAT

Note: All Inventories in man-quarters. Divide by 4 to obtain an average over the fiscal year.

Tape Characteristics:
IBM STANDARD LABEL TAPE; RECFM=FB; LRECL=69; BLKSIZE=17940;
DSN=MARINE.MOS; VOL=SER=000001; # OF BLOCKS ON THE TAPE=672

COLUMNS	DESCRIPTION
1-2	RECORD TYPE
3-4	YEARS OF COMMISSIONED SERVICE
5-6	PAY-GRADE
7-9	PRIMARY MOS
10	SEX
11-12	COMMISSIONING SOURCE
13	EDUCATION LEVEL
14	SERVICE COMPONENT
15-17	FIRST ADDITIONAL MOS
18-20	SECOND ADDITIONAL MOS
21	RACE
22	=1 if completed CLS school, 0 otherwise.
23	=1 if completed CLS extention, 0 otherwise.
24	=1 if completed ILS school, 0 otherwise.
25	=1 if completed ILS extension, 0 otherwise.
26	=1 if completed TLS school, 0 otherwise.
27	=1 if completed TLS extension, 0 otherwise.
28	=1 Did not complete any above, 0 otherwise.
29	BLANK
30-33	DATA for Fiscal YEAR 1977
34-37	DATA for Fiscal Year 1978
38-41	DATA for Fiscal Year 1979
42-45	DATA for Fiscal Year 1980
46-49	DATA for Fiscal Year 1981
50-53	DATA for Fiscal Year 1982
54-57	DATA for Fiscal Year 1983
58-61	DATA for Fiscal Year 1984
62-65	DATA for Fiscal Year 1985
66-69	DATA for Fiscal Year 1986

- RECORD TYPE (Note: *** means computed value; does not exist on file)

00=Inventory (man-QUARTER)

01=Retirement Loss

02=Release Loss

03=Discharge Loss

04=Resignation Loss

05=Other Loss

06=MOS change

07=All Strength Losses ***(All Voluntary + All Involuntary)

08=All Voluntary Losses

09=All Involuntary Losses

10=End of Obligated Service (EAS) Losses

11=Statutory Losses

12=Warrant to LDO flows

13=LDO to Warrant flows

YEARS OF COMMISSIONED SERVICE (YCS) (Note: *** means computed value; does not exist on file)

00=UNKNOWN
01=YCS 1
02=YCS 2
03=YCS 3
.
.
.
31=YCS 31
32=TOTAL ***

Pay Grade

00=UNKNOWN
01=WARRANT OFFICER W-1
02=CHIEF WARRANT OFFICER W-2
03=CHIEF WARRANT OFFICER W-3
04=CHIEF WARRANT OFFICER W-4
05=ALL WARRANT OFFICERS ***
06=LDO FIRST LIEUTENANT O-2
07=LDO FIRST LIEUTENANT O-2 FAILED SELECT
08=LDO CAPTAIN O-3
09=LDO CAPTAIN O-3 FAILED SELECT
10=LDO MAJOR O-4
11=LDO MAJOR O-4 FAILED SELECT
12=LDO LIEUTENANT COLONEL O-5
13=ALL LDOS ***
14=SECOND LIEUTENANT O-1
15=SECOND LIEUTENANT O-2
16=SECOND LIEUTENANT O-2 FAILED SELECT
17=CAPTAIN O-3
18=CAPTAIN O-3 FAILED SELECT
19=MAJOR O-4
20=MAJOR O-4 FAILED SELECT
21=LIEUTENANT COLONEL O-5
22=LIEUTENANT COLONEL O-5 FAILED SELECT
23=COLONEL O-6
24=ALL UNR OFFICERS ***
25=ALL MARINE CORPS OFFICERS ***

SEX

1=MALE
2=FEMALE
3=TOTAL ***

COMMISSIONING SOURCE

01=US NAVAL ACADEMY
02=PLATOON LEADER CLASS -AVIATION
03=PLATOON LEADER CLASS -GROUND
04=PLATOON LEADER CLASS -LAW
05=AVIATION OFFICER CANDIDATE
06=MARINE AVIATION CADET
07=OFFICER CANDIDATE COURSE -GROUND
08=OFFICER CANDIDATE COURSE -LAW
09=OFFICER CANDIDATE COURSE -WOMEN
10=ENLISTED COMMISSIONING PROGRAM
11=NROTC -SCHOLARSHIP
12=NROTC -GROUND COLLEGE
13=NROTC -AVIATION COLLEGE
14=NAVY ENLISTED SCIENTIFIC EDUATION PROGRAM (NESEP)
15=ALL OTHER SOURCES OF ENTRY INCLUDING RECALL
16=TOTAL ***

EDUCATION LEVEL

01=NON-COLLEGE GRADUATE
02=COLLEGE GRADUATE - 4 YEAR DEGREE OR PROFESSIONAL DEGREE
03=COLLEGE GRADUATE - MASTERS
04=COLLECT GRADUATE - DOCTORATE
05=ALL ***

SERVICE COMPONENT

01=REGULAR-DID NOT AUGMENT
02=REGULAR-DID AUGMENT
03=RESERVE
04=REGULAR-TOTAL ***
05=REGULAR + RESERVE ***

RACE

01=WHITE
02=BLACK
03=HISPANIC
04=OTHER
05=TOTAL ***

0101	001	001	BASIC PERSONNEL AND ADMINISTRATION OFFICER
0102	005	005	...0102=0180
0107	000	002	CIVIL AFFAIRS OFFICER
0108	005	005	...0108=0180
0130	004	004	...0130=0170
0160	003	003	POSTAL OFFICER
0170	004	004	ADMINISTRATIVE OFFICER
0180	005	005	ADJUTANT
0201	006	006	BASIC INTELLIGENCE OFFICER
0202	007	007	INTELLIGENCE OFFICER
0205	008	008	TACTICAL INTELLIGENCE OFFICER
0210	009	009	COUNTERINTELLIGENCE OFFICER
0240	000	010	IMAGERY INTERPRETATION OFFICER
0250	000	011	INTERROGATION-TRANSLATION OFFICER
0301	012	012	BASIC INFANTRY OFFICER
0302	013	013	INFANTRY OFFICER
0303	000	014	LIGHT-ARMORED VEHICLE OFFICER
0401	015	015	BASIC LOGISTICS OFFICER
0402	016	016	LOGISTICS OFFICER
0406	018	018	...0406=0430
0410	000	017	MAINTENANCE MANAGEMENT OFFICER
0430	018	018	EMBARKATION OFFICER
0450	018	018	...0450=0430
0801	019	019	BASIC FIELD ARTILLERY OFFICER
0802	020	020	FIELD ARTILLERY OFFICER
0803	021	021	SURVEY AND METEOROLOGICAL OFFICER
0805	021	021	...0805=0803
0840	022	022	NAVAL GUNFIRE PLANNER
0845	023	023	NAVAL GUNFIRE SPOTTER
1101	000	024	BASIC UTILITIES OFFICER
1120	025	025	UTILITIES OFFICER
1301	026	026	BASIC ENGINEER, CONSTRUCTION, AND EQUIP OFFICER
1302	027	027	ENGINEER OFFICER
1305	000	028	SHORE PARTY OFFICER
1310	029	029	ENGINEER EQUIPMENT OFFICER
1320	027	027	...1320=1302
1330	000	030	FACILITIES MANAGEMENT OFFICER
1360	031	031	CONSTRUCTION OFFICER
1390	032	032	BULK FUEL OFFICER
1401	033	033	BASIC MAPPING OFFICER
1402	034	034	MAPPING OFFICER
1501	035	035	BASIC PRINTING AND REPRODUCTION OFFICER
1502	036	036	REPRODUCTION OFFICER
1801	037	037	BASIC TANK AND AMPHIBIAN TRACTOR OFFICER
1802	038	038	TANK OFFICER
1803	039	039	ASSAULT AMPHIBIAN VEHICLE OFFICER
2002	000	040	...2002=2101
2010	041	041	...2010=2102
2020	043	043	...2020=2120
2025	044	044	...2025=2125
2040	047	047	...2040=2340
2045	046	046	...2045=2305
2101	000	040	BASIC ORDNANCE OFFICER
2102	041	041	ORDNANCE OFFICER
2110	042	042	ORDNANCE VEHICLE MAINTENANCE OFFICER

2120	043	043	WEAPONS REPAIR OFFICER
2125	044	044	ELECTRO-OPTIC INSTRUMENT REPAIR OFFICER
2170	052	052	...2170=2602
2301	045	045	BASIC AMMUNITION AND EOD OFFICER
2305	046	046	EOD OFFICER
2340	047	047	AMMUNITION OFFICER
2501	048	048	BASIC OPERATIONAL COMMUNICATION OFFICER
2502	049	049	COMMUNICATION OFFICER
2503	049	049	NEW TELECOMMUNICATIONS OPERATIONS - WARRANTS
2505	000	050	TELECOMMUNICATION SYSTEMS OFFICER
2510	052	052	...2510=2602
2601	051	051	BASIC SIGNALS INTELLIGENCE/EW OFFICER
2602	052	052	SIGNALS INTELLIGENCE/EW OFFICER
2801	000	053	BASIC DATA/COMMUNICATIONS OFFICER
2802	054	054	ELECTRONICS MAINTENANCE OFFICER (GROUND)
2803	055	055	...2803=2805
2805	055	055	DATA/COMMUNICATIONS OFFICER
2810	056	056	TELEPHONE SYSTEMS OFFICER
2820	000	057	CALIBRATION OFFICER
2830	058	058	GROUND RADAR MAINTENANCE OFFICER
3001	059	059	BASIC SUPPLY ADMINISTRATION AND OPERATIONS OFFICER
3002	060	060	GROUND SUPPLY OFFICER
3010	061	061	GROUND SUPPLY OPERATIONS OFFICER
3040	000	062	CONTRACTING OFFICER
3050	063	063	WAREHOUSING OFFICER
3060	064	064	AVIATION SUPPLY OFFICER
3070	065	065	AVIATION SUPPLY OPERATIONS OFFICER
3101	066	066	BASIC TRANSPORTATION OFFICER
3102	067	067	TRAFFIC MANAGEMENT OFFICER
3202	025	025	...3202=1120
3301	068	068	BASIC FOOD SERVICE OFFICER
3302	069	069	FOOD SERVICE OFFICER
3310	069	069	...3310=3302
3401	070	070	BASIC AUDITING, FINANCE, AND ACCOUNTING OFFICER
3402	071	071	DISBURSING OFFICER
3406	072	072	FINANCIAL ACCOUNTING OFFICER
3410	073	073	AUDITING OFFICER
3415	074	074	FINANCIAL MANAGEMENT OFFICER
3501	075	075	BASIC MOTOR TRANSPORT OFFICER
3502	076	076	MOTOR TRANSPORT OFFICER
3510	077	077	MOTOR TRANSPORT MAINTENANCE OFFICER
3800	079	079	...3800=4002
4001	078	078	BASIC DATA SYSTEMS OFFICER
4002	079	079	DATA SYSTEMS OFFICER
4003	080	080	...4003=4006
4005	081	081	...4005=4010
4006	080	080	DATA SYSTEMS OPERATIONS OFFICER
4010	081	081	DATA SYSTEMS SOFTWARE OFFICER
4101	082	082	BASIC MARINE CORPS EXCHANGE OFFICER
4130	083	083	MARINE CORPS EXCHANGE OFFICER
4301	084	084	BASIC PUBLIC AFFAIRS OFFICER
4302	085	085	PUBLIC AFFAIRS OFFICER
4330	086	086	HISTORICAL OFFICER
4401	087	087	STUDENT JUDGE ADVOCATE
4402	088	088	JUDGE ADVOCATE

4420	000	089	LEGAL SERVICES OFFICER
4430	090	090	LEGAL ADMINISTRATIVE OFFICER
4601	000	091	BASIC TRAINING AND AUDIOVISUAL SUPPORT OFFICER
4602	092	092	TRAINING AND AUDIOVISUAL SUPPORT OFFICER
4902	092	092	...4902=4602
4915	225	225	...4915=9925
5501	000	093	BASIC BAND OFFICER
5502	094	094	BAND OFFICER
5505	095	095	DRUM AND BUGLE CORPS OFFICER
5701	096	096	BASIC NUCLEAR, BIOLOGICAL, AND CHEMICAL DEFENSE OFFICER
5702	097	097	NUCLEAR, BIOLOGICAL, AND CHEMICAL DEFENSE OFFICER
5710	097	097	...5710=5702
5715	000	098	NUCLEAR AND CHEMICAL WEAPONS EMPLOYMENT OFFICER
5720	099	099	GROUND NUCLEAR WEAPONS ASSEMBLY OFFICER
5801	100	100	BASIC MILITARY POLICE AND CORRECTIONS OFFICER
5802	102	102	...5802=5804
5803	101	101	MILITARY POLICE OFFICER
5804	102	102	CORRECTIONS OFFICER
5805	103	103	CRIMINAL INVESTIGATIONS OFFICER
5901	104	104	BASIC ELECTRONICS MAINTENANCE OFFICER
5902	105	105	ELECTRONICS MAINTENANCE OFFICER (AVIATION)
5903	106	106	...5903=5907
5905	105	105	...5905=5902
5907	106	106	GROUND LAUNCHED MISSILE SYSTEM MAINTENANCE OFFICER
5910	107	107	AVIATION RADAR MAINTENANCE OFFICER
5920	108	108	...5920=5950
5950	108	108	AIR TRAFFIC CONTROL SYSTEMS MAINTENANCE OFFICER
5970	109	109	DATA SYSTEMS MAINTENANCE OFFICER
6001	110	110	BASIC AIRCRAFT MAINTENANCE OFFICER
6002	111	111	AIRCRAFT MAINTENANCE OFFICER
6004	112	112	AIRCRAFT MAINTENANCE ENGINEERING OFFICER
6005	113	113	AERONAUTICAL OFFICER
6007	114	114	FLIGHT EQUIPMENT OFFICER
6009	114	114	...6009=6007
6101	115	115	BASIC AIRCRAFT MAINTENANCE OFFICER
6102	116	116	AIRCRAFT MAINTENANCE OFFICER
6104	117	117	AIRCRAFT MAINTENANCE ENGINEERING OFFICER
6105	118	118	AERONAUTICAL OFFICER
6107	119	119	FLIGHT EQUIPMENT OFFICER
6202	121	121	...6202=6302
6301	120	120	BASIC AVIONICS OFFICER
6302	121	121	AVIONICS OFFICER
6401	122	122	BASIC AVIONICS OFFICER
6402	123	123	AVIONICS OFFICER
6501	124	124	BASIC AVIATION ORDNANCE OFFICER
6502	125	125	AVIATION ORDNANCE OFFICER
6505	126	126	MARINE WING WEAPONS UNIT OFFICER
6602	121	121	...6602=6302
6701	125	125	...6701=6502
6704	125	125	...6704=6502
6706	125	125	...6706=6502
6707	125	125	...6707=6502
6708	125	125	...6708=6502
6709	125	125	...6709=6502
6710	125	125	...6710=6502

6720	125	125	...6720=6502
6801	127	127	BASIC WEATHER SERVICE OFFICER
6802	128	128	WEATHER SERVICE OFFICER
7001	129	129	BASIC AIRFIELD SERVICES OFFICER
7002	130	130	AIRFIELD SERVICES OFFICER
7140	132	132	...7140=7204
7201	131	131	BASIC AIR CONTROL/ANTI-AIR WARFARE OFFICER
7203	132	132	...7203=7204
7204	132	132	ANTI-AIR WARFARE OFFICER
7207	000	133	FORWARD AIR CONTROLLER
7208	134	134	AIR SUPPORT CONTROL OFFICER
7210	135	135	AIR DEFENSE CONTROL OFFICER
7220	000	136	OA-4M FAC(A)/TAC(A)
7277	000	137	WEAPONS AND TACTICS INSTRUCTOR-AIR CONTROL
7301	138	138	BASIC AIR TRAFFIC CONTROL OFFICER
7320	139	139	AIR TRAFFIC CONTROL OFFICER
7330	140	140	RADAR APPROACH CONTROLLER
7380	141	141	AERIAL NAVIGATION OFFICER
7500	142	142	BASIC PILOT VMA
7501	143	143	PILOT VMA-A-4
7508	144	144	PILOT VMA-AV-8A/C
7509	145	145	PILOT VMA-AV-8B
7510	146	146	BASIC PILOT VMA (AW)
7511	147	147	PILOT VMA(AW) A-6
7518	148	148	...7518=7520
7519	148	148	...7519=7520
7520	148	148	BASIC PILOT VMFA (F-4)
7521	149	149	PILOT VMFA (F/A-18)
7522	150	150	PILOT VMFA F-4/J/S
7523	151	151	PILOT VMFA F/A-18
7528	152	152	...7528=7540
7529	152	152	...7529=7540
7530	148	148	...7530=7520
7531	148	148	...7530=7520
7532	148	148	...7532=7520
7540	152	152	BASIC PILOT VMAQ/VMFP
7541	153	153	...7541=7542
7542	153	153	PILOT VMAQ/EA-6A
7543	154	154	PILOT VMAQ/EA-6B
7545	155	155	PILOT VMFP/RF-4B
7550	156	156	BASIC PILOT VMGR
7551	000	157	PILOT VMGR C-9
7552	000	158	PILOT VMGR TC-4C
7553	160	160	...7553=7556
7554	187	187	...7554=7597
7555	000	159	PILOT UC-12B
7556	160	160	KC-130 CO-PILOT (T2P/T3P)
7557	161	161	KC-130 AIRCRAFT COMMANDER
7558	000	162	...7558=7559
7559	000	162	PILOT VMGR CT-39
7560	163	163	BASIC PILOT HMH/M/L/A
7561	164	164	...7561=7562
7562	164	164	PILOT HMM CH-46
7563	165	165	PILOT HML UH-1
7564	166	166	PILOT HMH CH-53

7565	167	167	PILOT HMA AH-1
7566	168	168	PILOT HMH CH-53E
7575	169	169	BASIC PILOT VMO
7576	170	170	PILOT VMO/OV-10
7577	000	171	WEAPONS AND TACTICS INSTRUCTOR
7580	172	172	FLIGHT OFFICER STUDENT
7581	173	173	BASIC NAVAL FLIGHT OFFICER
7582	178	178	...7582=7587
7583	174	174	BOMBADIER-NAVIGATOR A-6
7584	175	175	ELECTRONICS WARFARE OFFICER, EA-6A
7585	176	176	AIRBORNE RECONNAISSANCE OFFICER,RF-4B
7586	177	177	EW/AIRBORNE RECONNAISSANCE OFFICER, EA-6A/RF-4B
7587	178	178	AIRBORNE RADAR INTERCEPT OFFICER, F4N/J/S
7588	179	179	EW OFFICER, EA-6B
7590	000	180	LANDING SIGNAL OFFICER-TRAINEE
7591	000	181	NAVAL FLIGHT OFFICERVMAW
7592	000	182	PILOT VMAW
7593	000	183	LSO, PHASE I & II
7594	000	184	LSO, PHASE III
7595	000	185	TEST PILOT/FLIGHT TEST PROJECT OFFICER
7596	000	186	AVIATION SAFETY OFFICER
7597	187	187	BASIC ROTARY WING PILOT
7598	188	188	BASIC FIXED WING PILOT
7599	189	189	FLIGHT STUDENT
9602	000	190	EDUCATION OFFICER
9608	000	000	...???
9620	000	191	AERONAUTICAL ENGINEER
9622	000	192	CHEMICAL ENGINEER
9624	000	193	ELECTRONICS ENGINEER
9626	000	194	ORDNANCE SYSTEMS ENGINEER
9628	000	195	COMPUTER ENGINEER
9630	000	196	INDUSTRIAL ENGINEER
9632	000	197	NUCLEAR ENGINEER
9634	000	198	ELECTRONIC WARFARE SYSTEMS OFFICER
9636	000	199	COMMUNICATIONS ENGINEER
9638	000	000	...???
9640	000	200	MANAGEMENT OFFICER
9644	000	201	FINANCIAL MANAGEMENT SPECIALIST
9646	000	202	DATA SYSTEMS SPECIALIST
9648	000	203	MANAGEMENT, DATA SYSTEMS OFFICER
9650	000	204	OPERATIONS ANALYST
9652	000	205	DEFENSE SYSTEMS ANALYST
9654	000	000	...???
9656	000	206	SYSTEMS ACQUISITION MANAGEMENT OFFICER
9658	000	207	C3 SYSTEMS OFFICER
9660	000	208	COMMUNICATION MANAGER
9662	000	209	MATERIAL MANAGEMENT OFFICER
9670	000	210	STATISTICS OFFICER
9672	000	000	...???
9674	000	211	PUBLIC INFORMATION OFFICER
9676	000	212	INTERNATIONAL RELATIONS OFFICER
9678	000	213	HISTORIAN
9680	000	214	HUMAN RESOURCES MANAGEMENT SPECIALIST
9688	000	215	MASTER OF LAWS
9699	000	216	MARINE OFFICER INSTRUCTOR

9900	000	000	FUL...BASIC MARINE (ENLISTED)
9901	217	217	BASIC OFFICER
9903	218	218	GENERAL OFFICER
9904	219	219	COLONEL,LOGISTICS
9905	000	220	SPECIAL ASSIGNMENT OFFICER
9906	221	221	COLONEL, GROUND
9907	222	222	COLONEL, NAVAL AVIATOR/NAVAL FLIGHT OFFICER
9908	223	223	COLONEL, SUPPLY
9910	000	000	... BILLET DESIGNATOR
9913	000	224	SPECIAL SERVICES OFFICER
9914	225	225	COLONEL, JUDGE ADVOCATE
9920	226	226	...9920=9925
9925	226	226	RANGE OFFICER
9940	000	227	FORIEGN AREA OFFICER
9945	000	000	...???
9947	228	228	PSYCHOLOGICAL OPERATIONS OFFICER
9950	229	229	COMBAT ARTIST (OFFICER)
9952	230	230	SCUBA MARINE (OFFICER/ENLISTED)
9953	231	231	PARACHUTIST/SCUBA MARINE (OFFICER/ENLISTED)
9956	232	232	GROUND SAFETY SPECIALIST (OFFICER/ENLISTED)
9960	000	233	NAVAL AVIATION OBSERVER
9962	234	234	PARACHUTIST (OFFICER/ENLISTED)
9980	000	235	SURVEILLANCE SENSOR OFFICER
9981	000	236	TACTICAL DATA SYSTEMS SPECIALIST (OFFICER/ENLISTED)
9982	000	000	...???

APPENDIX B

EXAMPLE OF SUMMARY DATA FILE

This appendix contains an example of the raw data as organized in the summary data file. As this is a composite data file, the formatting peculiarities are noteworthy. Columnar designations are provided. The column descriptions are provided in Appendix A.

col.:

```
12345     10    15    20    25    30    35    40    45    50    55    60    65
01202101310742190016100100000000000100000000000000000000000000000000
01202101310722071005100100000000000000000000010000000000000000000000
01202101310121193016111010000000000000000000000100000000000000000000
01202101310322233000101100000000000000000000010000000000000000000000
01202101310131200000100000000000000000000001000000000000000000000000
01202101311112107100511000000000010000000000000000000000000000000000
01202101310732200000110000000000000000000000100000000000000000000000
01202101310722233083110000000000000000000000010000000000000000000000
01202101310332200074100000000000000000000000100000000000000000000000
01202101310722018000100100000000100000000000000000000000000000000000
01202101311143120102711001000000000000000000000000000100000000000000
01202101310321018005110100000000000000000000001000000000000000000000
01202101310232204203101110000000000000000000000000000010000000000000
01202101310532211000110000000000000000000010000000000000000000000000
01202101311113121223110010100000000001000000000000000000000000000000
01202101310722076000110100000000000000000000000000000000000000000001
01202101310321076016100100000000001000000000000000000000000000000000
01202201310331200005110100000000000000000000000000000000000100000000
01212101310322000001101000000000000000000000000000000000000000000001
01212101310321234000110000000000000000000000000000000000000000000001
01212101310332000000110100000000000000000000000100000000000000000000
01212101310332000000100001000000000000000000000010000000000000000000
01212101310322000000111000000000100000000000000000000000000000000000
01212101310331018089110010000000000000000000000000000100000000000000
01212101310322004000101100000000000000000000000010000000000000000000
01212101310722098000110100000000100000000000000000000000000000000000
01212101310332004000110001000000000000000000100000000000000000000000
01212101311112123400011010000000000001000000000000000000000000000000
01212101310332005000110100000000000000000000000000000100000000000000
01212101310222000001000100000000100000000000000000000000000000000000
```

APPENDIX C
LOSS RATE COMPUTER PROGRAMS

```
      PARAMETER (MGRP=6, NTBL=86, NOCC=26, NYCS=26)                    MCR00010
      INTEGER TYPE,YCS,PG,MOS,SEX,CSRCE,EDLVL,SVC,MOS1,MOS2,RACE       MCR00020
      INTEGER DATA(77:86)                                              MCR00030
C---  BY MOS GROUP                                                     MCR00040
      INTEGER TOT(5,MGRP,77:86)                                        MCR00050
      INTEGER GT(MGRP,77:86)                                           MCR00060
C---  BY OCC GROUP                                                     MCR00070
      INTEGER TOTOCC(5,NOCC,77:86)                                     MCR00080
      INTEGER GTOCC(NOCC,77:86)                                        MCR00090
C---  BY YCS (1-NYCS)                                                  MCR00100
      INTEGER TOTYCS(5,NYCS,77:86)                                     MCR00110
      INTEGER GTYCS(NYCS,77:86)                                        MCR00120
C---                                                                   MCR00130
      CHARACTER*7 LBTYPE(5)                                            MCR00140
      CHARACTER*7 CITLS                                                MCR00150
      CHARACTER*8 MOSGRP(MGRP)                                         MCR00160
      INTEGER MOSTBL(2,NTBL)                                           MCR00170
      INTEGER OCCTBL(3,NOCC)                                           MCR00180
      INTEGER YCSTBL(3,4)                                              MCR00190
      REAL WTBL(78:86)                                                 MCR00200
      DATA GT/60*0./, TOT/300*0./                                      MCR00210
      DATA GTOCC/260*0./, TOTOCC/1300*0./                              MCR00220
      DATA GTYCS/260*0./, TOTYCS/1300*0./                              MCR00230
      DATA LBTYPE /'RETIRE','RELEASE','DISCH','RESIGN','OTHER'/        MCR00240
      DATA MOSGRP /'P(+)', 'F-18', 'GRD(_)',                           MCR00250
     * 'MTO/$', 'G', 'ALLELSE'/

      DATA MOSTBL /                                                    MCR00270
     * 143,2, 147,2, 150,2, 153,2, 154,2, 155,2, 170,2,                MCR00320
     * 142,1, 146,1, 148,1, 152,1, 156,1, 163,1, 169,1,                MCR00310
     * 157,2, 158,2, 159,2, 160,2, 161,2, 162,2, 164,2,
     * 167,2, 168,2, 178,2,                                            MCR00340
     * 173,2, 174,2, 175,2, 176,2, 177,2, 178,2, 144,2, 145,2,         MCR00350
     * 165,2, 7,2,  74,2,
     * 172,1, 187,1, 188,1, 189,1,
     * 5,4, 13,4, 20,4, 27,4, 38,4, 52,4, 132,4, 135,4, 139,4, 166,4,
     * 16,4, 39,4, 49,4, 60,4, 64,4, 87,4, 88,4, 101,4, 134,4,
     * 79,2, 85,2, 111,2, 116,2,
     * 1,1, 6,1, 12,1, 15,1, 19,1, 26,1, 37,1, 48,1, 51,1, 59,1,       MCR00290
     * 70,1, 75,1, 78,1, 84,1, 100,1, 110,1, 115,1, 131,1, 138,1,
     * 217,1,
     * 149,3, 151,3,
     * 71,5,  76,5    /
C                                                                      MCR00390
      DATA OCCTBL /1,5,1, 6,11,2, 12,14,3, 15,18,4, 19,23,5,           MCR00400
     * 26,32,6, 37,39,7,                                               MCR00410
     * 48,50,8, 51,52,9, 59,65,10,                                     MCR00420
     * 70,74,11, 75,77,12, 78,81,13, 84,86,14,                         MCR00430
     * 87,90,15, 100,103,16,                                           MCR00440
     * 110,114,17,                                                     MCR00450
     * 131,137,18, 138,141,19, 142,159,20,                             MCR00460
     * 160,161,21, 163,168,22, 169,170,23, 172,173,24,                 MCR00470
     * 174,179,25, 187,189,26  /                                       MCR00480
C---                                                                   MCR00490
      DATA YCSTBL /1,6,32, 7,12,33, 13,20,34, 21,31,35/                MCR00500
      DATA WTBL/1.,1.,1.,1.,1., 3.,5.,7.,9/                            MCR00510
      TOTW=0                                                           MCR00520
      DO 5 I=77,86                                                     MCR00530
       TOTW=TOTW+WTBL(I)                                               MCR00540
  5   CONTINUE                                                         MCR00550
      DO 7 I=77,86                                                     MCR00560
```

```
      WTBL(I)=WTBL(I)/TOTW                                              MCR00570
    7 CONTINUE                                                          MCR00580
C ---                                                                   MCR00590
      NREC=0                                                            MCR00600
      DO 50 KK=1,999999                                                 MCR00610
       READ(1,100,END=999) TYPE,YCS,PG,MOS,SEX,CSRCE,EDLVL,SVC,         MCR00620
     *  MOS1,MOS2,RACE,CITLS,DATA                                       MCR00630
       NREC=NREC+1                                                      MCR00640
       IF (PG.GE.1 .AND. PG.LE.12) GO TO 50                             MCR00650
C---  MOS GROUP CLASSIFICATIONS                                         MCR00660
       IF(TYPE.GE.1 .AND. TYPE.LE.5 .AND. MOS.LE.MOS) THEN              MCR00670
             IMOS=MOSGET(MOS,MOSTBL,NTBL,MGRP)                          MCR00680
             DO 10 I=77,86                                              MCR00690
              TOT(TYPE,IMOS,I)=TOT(TYPE,IMOS,I) + DATA(I)               MCR00700
   10        CONTINUE                                                   MCR00710
       ENDIF                                                            MCR00720
C---   TOTAL STRENGTH                                                   MCR00730
       IF(TYPE.EQ.0 .AND. YCS.LE.MOS) THEN                              MCR00740
             IMOS=MOSGET(MOS,MOSTBL,NTBL,MGRP)                          MCR00750
             DO 20 I=77,86                                              MCR00760
              GT(IMOS,I)=GT(IMOS,I) + DATA(I)                           MCR00770
   20        CONTINUE                                                   MCR00780
       ENDIF                                                            MCR00790
C --- OCC GROUP CLASSIFICATIONS                                         MCR00800
       IF(TYPE.GE.1 .AND. TYPE.LE.5) THEN                               MCR00810
             IOCC=NOCGET(MOS,OCCTBL,NOCC)                               MCR00820
             IF(IOCC .GT. 0) THEN                                       MCR00830
              DO 25 I=77,86                                             MCR00840
               TOTOCC(TYPE,IOCC,I)=TOTOCC(TYPE,IOCC,I) + DATA(I)        MCR00850
   25          CONTINUE                                                 MCR00860
             ENDIF                                                      MCR00870
       ENDIF                                                            MCR00880
C ---  TOTAL STRENGTH FOR OCC GROUPS                                    MCR00890
       IF(TYPE.EQ.0) THEN                                               MCR00900
             IOCC=NOCGET(MOS,OCCTBL,NOCC)                               MCR00910
             IF(IOCC .GT. 0) THEN                                       MCR00920
              DO 28 I=77,86                                             MCR00930
               GTOCC(IOCC,I)=GTOCC(IOCC,I) + DATA(I)                    MCR00940
   28          CONTINUE                                                 MCR00950
             ENDIF                                                      MCR00960
       ENDIF                                                            MCR00970
C --- YCS GROUP CLASSIFICATIONS                                         MCR00980
       IF(TYPE.GE.1 .AND. TYPE.LE.5 .AND. YCS.LE.NYCS) THEN             MCR00990
                DO 32 I=77,86                                           MCR01000
                 TOTYCS(TYPE,YCS,I)=TOTYCS(TYPE,YCS,I) + DATA(I)        MCR01010
   32           CONTINUE                                                MCR01020
       ENDIF                                                            MCR01030
C ---  TOTAL STRENGTH FOR YCS GROUPS                                    MCR01040
       IF(TYPE.EQ.0 .AND. YCS.LE.NYCS) THEN                             MCR01050
                DO 42 I=77,86                                           MCR01060
                 GTYCS(YCS,I)=GTYCS(YCS,I) + DATA(I)                    MCR01070
   42           CONTINUE                                                MCR01080
       ENDIF                                                            MCR01090
   50 CONTINUE                                                          MCR01100
C---                                                                    MCR01110
  999 CONTINUE                                                          MCR01120
C --- MOS GROUP RATES                                                   MCR01130
      WRITE(6,111)                                                      MCR01140
      CALL RATES(TOT,GT,MGRP,2,WTBL)                                    MCR01150
C ---                                                                   MCR01160
C --- OCC GROUP RATES                                                   MCR01170
      WRITE(6,121)                                                      MCR01180
      CALL RATES(TOTOCC,GTOCC,NOCC,3,WTBL)                              MCR01190
C ---                                                                   MCR01200
C --- YCS GROUP RATES                                                   MCR01210
      WRITE(6,122)                                                      MCR01220
      CALL RATES(TOTYCS,GTYCS,NYCS,4,WTBL)                              MCR01230
C ---                                                                   MCR01240
      WRITE(6,*) '***** TOTAL RECORDS=',NREC                            MCR01250
C---                                                                    MCR01260
```

```
  100 FORMAT(3I2,I3,I1,I2,2I1,2I3,I1,A7, 1X, 10I4)                    MCR01270
  111 FORMAT(' GROUP TYPE  ---MOSGRP DATA 77 TO 86 -------------',/)  MCR01280
  121 FORMAT(' GROUP TYPE  ---OCCGRP DATA 77 TO 86 -------------',/)  MCR01290
  122 FORMAT(' GROUP TYPE  ---YCSGRP DATA 77 TO 86 -------------',/)  MCR01300
      END                                                             MCR01310
C ---                                                                 MCR01320
      FUNCTION MOSGET(MOS,MOSTBL,NTBL,MGRP)                           MCR01330
      INTEGER MOSTBL(2,NTBL)                                          MCR01340
      DO 10 I=1,NTBL                                                  MCR01350
       IF(MOS.EQ.MOSTBL(1,I)) THEN                                    MCR01360
            MOSGET=MOSTBL(2,I)                                        MCR01370
            RETURN                                                    MCR01380
       ENDIF                                                          MCR01390
   10  CONTINUE                                                       MCR01400
      MOSGET=MGRP                                                     MCR01410
      END                                                             MCR01420
C                                                                     MCR01430
      FUNCTION NOCGET(MOS,OCCTBL,NOCC)                                MCR01440
      INTEGER OCCTBL(3,NOCC)                                          MCR01450
      DO 10 I=1,NOCC                                                  MCR01460
       IF(MOS.GE.OCCTBL(1,I) .AND. MOS.LE.OCCTBL(2,I)) THEN           MCR01470
            NOCGET=OCCTBL(3,I)                                        MCR01480
            RETURN                                                    MCR01490
       ENDIF                                                          MCR01500
   10  CONTINUE                                                       MCR01510
      NOCGET=0                                                        MCR01520
      END                                                             MCR01530
C                                                                     MCR01540
      FUNCTION NYCGET(YCS,YCSTBL,NY)                                  MCR01550
      INTEGER YCSTBL(3,NY), YCS                                       MCR01560
      DO 10 I=1,NY                                                    MCR01570
       IF(YCS.GE.YCSTBL(1,I) .AND. YCS.LE.YCSTBL(2,I)) THEN           MCR01580
            NYCGET=YCSTBL(3,I)                                        MCR01590
            RETURN                                                    MCR01600
       ENDIF                                                          MCR01610
   10  CONTINUE                                                       MCR01620
      NYCGET=0                                                        MCR01630
      END                                                             MCR01640
C                                                                     MCR01650
      SUBROUTINE RATES(TOT,GT,M,IFILE,WTBL)                           MCR01660
C ---  COMPUTE RATES                                                  MCR01670
      INTEGER TOT(5,M,77:86)                                          MCR01680
      INTEGER GT(M,77:86)                                             MCR01690
      REAL RATE(5,500,78:86), SUM(78:86)                              MCR01700
      REAL RATES(5,500)                                               MCR01710
      REAL WRATES(5,500)                                              MCR01720
      REAL WTBL(78:86)                                                MCR01730
      IF=IFILE+10                                                     MCR01740
      DO 200 J=1,M                                                    MCR01750
       DO 210 I=1,5                                                   MCR01760
        WRITE(6,105) J,I,(TOT(I,J,K),K=77,86)                         MCR01770
        WRITE(6,105) J,I,(GT(J,K),K=77,86)                            MCR01780
        WNUM=0                                                        MCR01790
        WDNUM=0                                                       MCR01800
        DO 220 K=78,86                                                MCR01810
         SUM(K)=.125*(GT(J,K-1)+GT(J,K))                              MCR01820
         IF(SUM(K).NE.0) THEN                                         MCR01830
                  T=TOT(I,J,K)/SUM(K)                                 MCR01840
                  RATE(I,J,K)=AMIN1(1., T)                            MCR01850
         ELSE                                                         MCR01860
                  RATE(I,J,K)=0.                                      MCR01870
         ENDIF                                                        MCR01880
         WNUM=WNUM+WTBL(K)*TOT(I,J,K)                                 MCR01890
         WDNUM=WDNUM+WTBL(K)*SUM(K)                                   MCR01900
   220   CONTINUE                                                     MCR01910
C ---    5-YEAR RATE                                                  MCR01920
        ANUM=0                                                        MCR01930
        DNUM=0                                                        MCR01940
        DO 222 K=82,86                                                MCR01950
         ANUM=ANUM+TOT(I,J,K)                                         MCR01960
```

```
       DNUM=DNUM+SUM(K)                                                    MCR01970
  222  CONTINUE                                                            MCR01980
       IF(DNUM.NE.0) THEN                                                  MCR01990
                     T=ANUM/DNUM                                           MCR02000
                     RATE5(I,J)=AMIN1(1., T)                               MCR02010
       ELSE                                                                MCR02020
                     RATE5(I,J)=0                                          MCR02030
       ENDIF                                                               MCR02040
       IF(WDNUM.NE.0) THEN                                                 MCR02050
                     T=WNUM/WDNUM                                          MCR02060
                     WRATE5(I,J)=AMIN1(1., T)                              MCR02070
       ELSE                                                                MCR02080
                     WRATE5(I,J)=0                                         MCR02090
       ENDIF                                                               MCR02100
       WRITE(6,106) J,I,(RATE(I,J,K),K=78,86), RATE5(I,J),WRATE5(I,J)      MCR02110
  210  CONTINUE                                                            MCR02120
       DNUM=.2*DNUM                                                        MCR02130
       WRITE(IF,115) J, (SUM(K), K=78,86), DNUM, WDNUM                     MCR02140
  200 CONTINUE                                                             MCR02150
C --- WRITE DISK FILE WITH RATES ONE MATRIX FOR EACH YR                    MCR02160
      WRITE(IFILE,112) ((K,J,(RATE(I,J,K),I=1,5),J=1,M),K=78,86)           MCR02170
      WRITE(IFILE,112) ((K,J,(RATE5(I,J),I=1,5),J=1,M),K=98,98)            MCR02180
      WRITE(IFILE,112) ((K,J,(WRATE5(I,J),I=1,5),J=1,M),K=99,99)           MCR02190
  105 FORMAT(2I5, 10I6)                                                    MCR02200
  106 FORMAT(2I5, 10X, 11F6.3)                                             MCR02210
  112 FORMAT(2I5,  5F7.3)                                                  MCR02220
  115 FORMAT(I5,  11F8.0)                                                  MCR02230
      END                                                                  MCR02240
```

APPENDIX D
EXAMPLE LOSS RATE MATRIX FILE

78	1	0.001	0.014	0.000	0.003	0.000
78	2	0.000	0.004	0.006	0.006	0.002
78	3	0.002	0.152	0.004	0.008	0.002
78	4	0.002	0.124	0.002	0.030	0.003
78	5	0.000	0.088	0.003	0.038	0.003
78	6	0.003	0.130	0.012	0.094	0.002
78	7	0.001	0.038	0.020	0.087	0.001
78	8	0.001	0.055	0.004	0.053	0.004
78	9	0.004	0.021	0.000	0.049	0.008
78	10	0.003	0.012	0.000	0.049	0.002
78	11	0.003	0.004	0.000	0.023	0.004
78	12	0.087	0.009	0.045	0.021	0.003
78	13	0.076	0.000	0.012	0.015	0.009
78	14	0.015	0.000	0.000	0.000	0.004
78	15	0.023	0.000	0.000	0.003	0.000
78	16	0.020	0.000	0.000	0.000	0.003
78	17	0.021	0.000	0.000	0.000	0.003
78	18	0.009	0.000	0.000	0.000	0.000
78	19	0.048	0.003	0.000	0.000	0.000
78	20	0.295	0.003	0.000	0.000	0.000
78	21	0.294	0.000	0.000	0.000	0.005
78	22	0.140	0.000	0.000	0.000	0.000
78	23	0.173	0.000	0.000	0.000	0.000
78	24	0.063	0.005	0.000	0.000	0.000
78	25	0.180	0.000	0.000	0.000	0.000
78	26	0.328	0.000	0.000	0.000	0.000
79	1	0.000	0.012	0.001	0.004	0.001
79	2	0.001	0.005	0.003	0.002	0.001
79	3	0.001	0.071	0.004	0.005	0.001
79	4	0.002	0.188	0.004	0.033	0.004
79	5	0.001	0.091	0.002	0.058	0.003
79	6	0.001	0.090	0.009	0.055	0.008
79	7	0.004	0.048	0.000	0.063	0.003
79	8	0.003	0.049	0.005	0.083	0.000
79	9	0.000	0.032	0.002	0.079	0.003
79	10	0.001	0.015	0.001	0.043	0.003
79	11	0.013	0.003	0.000	0.044	0.005
79	12	0.003	0.006	0.049	0.023	0.000
79	13	0.074	0.008	0.064	0.033	0.008
79	14	0.082	0.006	0.000	0.013	0.000
79	15	0.026	0.000	0.000	0.004	0.004
79	16	0.016	0.003	0.000	0.003	0.000
79	17	0.020	0.000	0.000	0.006	0.000
79	18	0.021	0.000	0.000	0.000	0.000
79	19	0.075	0.000	0.000	0.000	0.000
79	20	0.307	0.003	0.000	0.000	0.003
79	21	0.206	0.005	0.000	0.000	0.000
79	22	0.144	0.000	0.000	0.000	0.006
79	23	0.129	0.000	0.000	0.000	0.000
79	24	0.114	0.000	0.000	0.000	0.007
79	25	0.113	0.006	0.000	0.006	0.000
79	26	0.341	0.000	0.000	0.000	0.011
80	1	0.000	0.013	0.000	0.008	0.000
80	2	0.001	0.005	0.001	0.006	0.001
80	3	0.000	0.028	0.004	0.002	0.001
80	4	0.001	0.157	0.003	0.032	0.003
80	5	0.002	0.052	0.003	0.066	0.003
80	6	0.002	0.078	0.007	0.072	0.003
80	7	0.000	0.068	0.010	0.076	0.000
80	8	0.002	0.027	0.002	0.069	0.005
80	9	0.002	0.019	0.005	0.051	0.004

80	10	0.000	0.010	0.002	0.034	0.000
80	11	0.006	0.016	0.000	0.019	0.003
80	12	0.020	0.005	0.068	0.022	0.003
80	13	0.019	0.002	0.090	0.019	0.005
80	14	0.043	0.009	0.000	0.009	0.004
80	15	0.040	0.000	0.000	0.000	0.003
80	16	0.027	0.000	0.000	0.004	0.015
80	17	0.020	0.000	0.000	0.000	0.003
80	18	0.020	0.000	0.000	0.000	0.000
80	19	0.048	0.008	0.000	0.000	0.000
80	20	0.307	0.000	0.000	0.003	0.000
80	21	0.210	0.010	0.000	0.000	0.000
80	22	0.132	0.000	0.000	0.000	0.006
80	23	0.144	0.000	0.000	0.000	0.014
80	24	0.138	0.000	0.000	0.000	0.000
80	25	0.123	0.000	0.000	0.000	0.008
80	26	0.212	0.000	0.000	0.000	0.000
81	1	0.000	0.008	0.001	0.006	0.000
81	2	0.000	0.005	0.001	0.006	0.001
81	3	0.000	0.018	0.003	0.002	0.001
81	4	0.003	0.144	0.003	0.025	0.005
81	5	0.001	0.059	0.002	0.058	0.004
81	6	0.003	0.075	0.017	0.079	0.006
81	7	0.001	0.053	0.006	0.067	0.005
81	8	0.000	0.017	0.001	0.048	0.006
81	9	0.003	0.015	0.000	0.044	0.000
81	10	0.006	0.008	0.002	0.034	0.002
81	11	0.005	0.013	0.011	0.018	0.002
81	12	0.008	0.002	0.084	0.031	0.002
81	13	0.033	0.006	0.058	0.020	0.002
81	14	0.008	0.000	0.000	0.008	0.002
81	15	0.039	0.002	0.000	0.002	0.002
81	16	0.034	0.000	0.000	0.003	0.007
81	17	0.032	0.000	0.000	0.000	0.000
81	18	0.017	0.000	0.000	0.000	0.003
81	19	0.030	0.000	0.000	0.000	0.000
81	20	0.310	0.000	0.000	0.000	0.000
81	21	0.213	0.005	0.000	0.000	0.000
81	22	0.094	0.000	0.000	0.000	0.000
81	23	0.146	0.000	0.000	0.000	0.007
81	24	0.080	0.008	0.000	0.000	0.000
81	25	0.132	0.000	0.000	0.000	0.019
81	26	0.271	0.000	0.000	0.000	0.000
82	1	0.000	0.009	0.003	0.008	0.000
82	2	0.000	0.011	0.006	0.006	0.001
82	3	0.001	0.007	0.001	0.002	0.000
82	4	0.001	0.139	0.003	0.016	0.003
82	5	0.001	0.052	0.001	0.048	0.002
82	6	0.002	0.026	0.011	0.047	0.002
82	7	0.001	0.042	0.014	0.056	0.001
82	8	0.001	0.019	0.001	0.020	0.001
82	9	0.003	0.008	0.003	0.023	0.000
82	10	0.002	0.005	0.004	0.023	0.000
82	11	0.006	0.008	0.092	0.004	0.000
82	12	0.008	0.004	0.079	0.014	0.002
82	13	0.019	0.002	0.017	0.011	0.000
82	14	0.019	0.000	0.000	0.004	0.000
82	15	0.006	0.000	0.000	0.000	0.002
82	16	0.009	0.005	0.000	0.000	0.002
82	17	0.028	0.004	0.000	0.004	0.004
82	18	0.033	0.004	0.000	0.000	0.000
82	19	0.039	0.000	0.000	0.000	0.000
82	20	0.290	0.000	0.000	0.000	0.000
82	21	0.239	0.005	0.000	0.000	0.000
82	22	0.181	0.000	0.000	0.000	0.000
82	23	0.165	0.000	0.000	0.000	0.000
82	24	0.144	0.000	0.000	0.000	0.000
82	25	0.182	0.000	0.000	0.000	0.000
82	26	0.231	0.000	0.000	0.000	0.000
83	1	0.001	0.007	0.003	0.011	0.001

83	2	0.001	0.001	0.006	0.006	0.001
83	3	0.001	0.003	0.001	0.003	0.003
83	4	0.001	0.122	0.002	0.013	0.003
83	5	0.000	0.074	0.004	0.036	0.001
83	6	0.000	0.034	0.008	0.032	0.003
83	7	0.001	0.064	0.010	0.029	0.001
83	8	0.000	0.033	0.000	0.019	0.001
83	9	0.001	0.012	0.001	0.025	0.000
83	10	0.003	0.003	0.003	0.011	0.002
83	11	0.004	0.004	0.072	0.007	0.000
83	12	0.000	0.000	0.024	0.007	0.002
83	13	0.013	0.000	0.011	0.009	0.000
83	14	0.006	0.002	0.000	0.002	0.000
83	15	0.008	0.000	0.000	0.000	0.000
83	16	0.006	0.002	0.000	0.004	0.002
83	17	0.012	0.000	0.002	0.002	0.000
83	18	0.040	0.000	0.000	0.000	0.000
83	19	0.017	0.000	0.000	0.004	0.004
83	20	0.221	0.000	0.000	0.000	0.004
83	21	0.107	0.000	0.005	0.000	0.000
83	22	0.123	0.000	0.000	0.000	0.000
83	23	0.126	0.000	0.000	0.000	0.000
83	24	0.074	0.000	0.000	0.000	0.008
83	25	0.137	0.000	0.000	0.000	0.009
83	26	0.131	0.000	0.000	0.000	0.000
84	1	0.001	0.013	0.011	0.005	0.000
84	2	0.000	0.003	0.010	0.006	0.001
84	3	0.000	0.001	0.001	0.005	0.004
84	4	0.001	0.183	0.002	0.008	0.005
84	5	0.000	0.083	0.003	0.035	0.003
84	6	0.000	0.051	0.016	0.036	0.003
84	7	0.001	0.080	0.010	0.049	0.005
84	8	0.001	0.039	0.004	0.035	0.004
84	9	0.000	0.020	0.001	0.033	0.002
84	10	0.000	0.011	0.003	0.011	0.006
84	11	0.003	0.003	0.055	0.014	0.003
84	12	0.004	0.000	0.078	0.012	0.004
84	13	0.002	0.007	0.005	0.005	0.000
84	14	0.002	0.007	0.002	0.002	0.004
84	15	0.010	0.004	0.002	0.002	0.002
84	16	0.015	0.000	0.000	0.004	0.002
84	17	0.011	0.002	0.000	0.000	0.000
84	18	0.054	0.002	0.000	0.000	0.002
84	19	0.049	0.000	0.000	0.000	0.000
84	20	0.179	0.004	0.000	0.000	0.000
84	21	0.131	0.000	0.000	0.000	0.000
84	22	0.073	0.000	0.000	0.000	0.006
84	23	0.141	0.000	0.000	0.000	0.006
84	24	0.104	0.000	0.000	0.000	0.008
84	25	0.170	0.000	0.000	0.000	0.009
84	26	0.346	0.000	0.000	0.000	0.000
85	1	0.000	0.001	0.004	0.068	0.000
85	2	0.000	0.003	0.006	0.002	0.001
85	3	0.001	0.002	0.004	0.012	0.001
85	4	0.002	0.316	0.005	0.023	0.003
85	5	0.003	0.097	0.006	0.042	0.002
85	6	0.001	0.049	0.018	0.048	0.006
85	7	0.002	0.125	0.010	0.052	0.002
85	8	0.002	0.040	0.005	0.048	0.003
85	9	0.001	0.026	0.003	0.033	0.002
85	10	0.001	0.012	0.004	0.022	0.003
85	11	0.007	0.007	0.054	0.013	0.000
85	12	0.013	0.009	0.122	0.013	0.002
85	13	0.006	0.004	0.004	0.011	0.002
85	14	0.002	0.012	0.000	0.007	0.000
85	15	0.004	0.000	0.000	0.002	0.000
85	16	0.008	0.002	0.000	0.000	0.002
85	17	0.019	0.002	0.004	0.000	0.000
85	18	0.013	0.002	0.000	0.000	0.000
85	19	0.044	0.000	0.000	0.000	0.000

85	20	0.232	0.007	0.000	0.000	0.004
85	21	0.178	0.000	0.000	0.000	0.000
85	22	0.094	0.000	0.000	0.000	0.000
85	23	0.120	0.000	0.000	0.000	0.000
85	24	0.092	0.000	0.000	0.000	0.000
85	25	0.092	0.000	0.000	0.000	0.000
85	26	0.217	0.000	0.000	0.000	0.000
86	1	0.000	0.001	0.003	0.005	0.000
86	2	0.001	0.002	0.002	0.004	0.002
86	3	0.001	0.003	0.004	0.003	0.001
86	4	0.001	0.288	0.004	0.014	0.001
86	5	0.000	0.092	0.006	0.035	0.003
86	6	0.001	0.076	0.016	0.056	0.001
86	7	0.004	0.137	0.008	0.075	0.005
86	8	0.001	0.045	0.002	0.059	0.002
86	9	0.002	0.023	0.001	0.053	0.001
86	10	0.004	0.007	0.002	0.024	0.002
86	11	0.003	0.008	0.015	0.036	0.001
86	12	0.008	0.007	0.066	0.025	0.002
86	13	0.017	0.000	0.008	0.017	0.000
86	14	0.002	0.002	0.000	0.017	0.002
86	15	0.007	0.000	0.002	0.005	0.002
86	16	0.011	0.005	0.000	0.005	0.000
86	17	0.012	0.002	0.000	0.000	0.000
86	18	0.013	0.004	0.000	0.002	0.000
86	19	0.036	0.000	0.000	0.000	0.000
86	20	0.249	0.003	0.000	0.000	0.000
86	21	0.184	0.000	0.000	0.005	0.000
86	22	0.226	0.000	0.000	0.000	0.000
86	23	0.100	0.000	0.000	0.000	0.000
86	24	0.077	0.000	0.000	0.000	0.007
86	25	0.039	0.000	0.000	0.000	0.000
86	26	0.220	0.000	0.000	0.000	0.000
98	1	0.000	0.007	0.005	0.019	0.000
98	2	0.000	0.004	0.006	0.005	0.001
98	3	0.001	0.003	0.002	0.005	0.001
98	4	0.001	0.208	0.003	0.015	0.003
98	5	0.001	0.079	0.004	0.039	0.002
98	6	0.001	0.047	0.014	0.044	0.003
98	7	0.002	0.090	0.010	0.052	0.003
98	8	0.001	0.035	0.003	0.037	0.003
98	9	0.002	0.018	0.002	0.034	0.001
98	10	0.002	0.008	0.003	0.018	0.003
98	11	0.005	0.006	0.054	0.016	0.001
98	12	0.007	0.004	0.076	0.015	0.002
98	13	0.012	0.002	0.009	0.011	0.000
98	14	0.006	0.004	0.000	0.006	0.001
98	15	0.007	0.001	0.001	0.002	0.001
98	16	0.010	0.003	0.000	0.003	0.002
98	17	0.016	0.002	0.001	0.001	0.000
98	18	0.029	0.003	0.000	0.001	0.001
98	19	0.038	0.000	0.000	0.001	0.001
98	20	0.239	0.003	0.000	0.000	0.001
98	21	0.168	0.001	0.001	0.001	0.000
98	22	0.136	0.000	0.000	0.000	0.001
98	23	0.130	0.000	0.000	0.000	0.001
98	24	0.098	0.000	0.000	0.000	0.005
98	25	0.121	0.000	0.000	0.000	0.004
98	26	0.231	0.000	0.000	0.000	0.000
99	1	0.000	0.006	0.005	0.021	0.000
99	2	0.000	0.003	0.005	0.004	0.001
99	3	0.001	0.012	0.003	0.006	0.001
99	4	0.001	0.234	0.003	0.018	0.003
99	5	0.001	0.085	0.005	0.040	0.002
99	6	0.001	0.060	0.015	0.050	0.003
99	7	0.002	0.103	0.009	0.058	0.003
99	8	0.001	0.039	0.003	0.047	0.003
99	9	0.002	0.022	0.002	0.041	0.002
99	10	0.002	0.009	0.003	0.022	0.003
99	11	0.005	0.007	0.036	0.022	0.001

99	12	0.011	0.005	0.078	0.018	0.002
99	13	0.016	0.003	0.014	0.013	0.001
99	14	0.007	0.005	0.000	0.009	0.002
99	15	0.010	0.001	0.001	0.003	0.001
99	16	0.012	0.002	0.000	0.003	0.002
99	17	0.015	0.002	0.001	0.000	0.000
99	18	0.023	0.003	0.000	0.001	0.001
99	19	0.041	0.000	0.000	0.000	0.000
99	20	0.244	0.003	0.000	0.000	0.001
99	21	0.174	0.001	0.001	0.002	0.000
99	22	0.137	0.000	0.000	0.000	0.002
99	23	0.124	0.000	0.000	0.000	0.002
99	24	0.089	0.001	0.000	0.000	0.005
99	25	0.102	0.000	0.000	0.000	0.003
99	26	0.247	0.000	0.000	0.000	0.001

APPENDIX E

METHOD TESTING

To determine the most useful clustering criterion for this project, a test was constructed to simulate loss rate data and compare the clustering solutions of two alternative methods. The clustering criterion considered were the average linkage between groups and the average linkage within groups.

Twelve sets of random numbers were generated to simulate officer attrition loss rates. The data are provided in Table E.1.

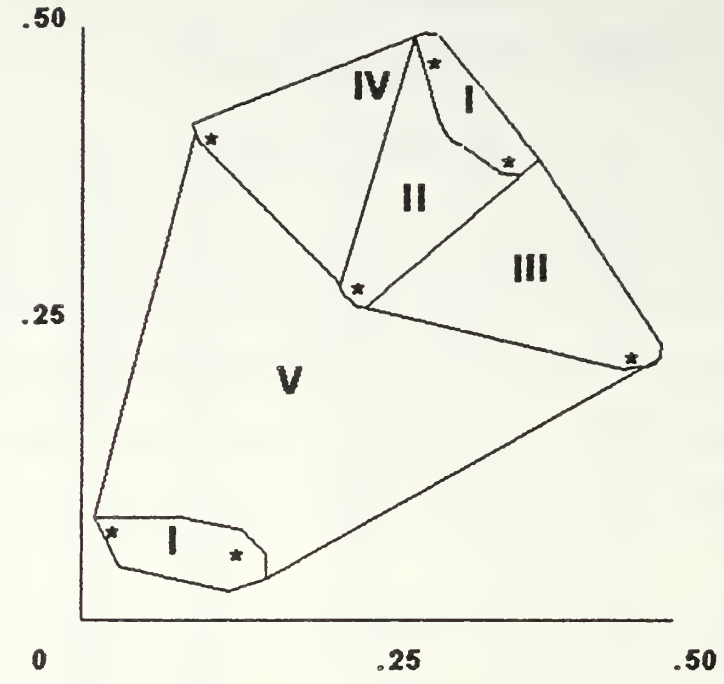
The 12 data sets were clustered according to the clustering criterion of each of the two candidate methods. The SPSSx CLUSTER procedure as offered by Norusis [Ref. 20:pp. 184-187] allows for both methods of clustering criterion to be specified by subcommand. The alternatives for each test were plotted and compared.

As anticipated, most of the tests (8 of 12) demonstrated little or no significant difference in the developed aggregation hierarchy. An example of such test similarity is provided in Figure E.1. Four test comparisons did, however, provide distinct, interesting, and consistent clustering trends. Examples of tests of interest are provided in Figures E.2 and E.3.

TABLE E.1

METHOD SIMULATION DATA SET

01	.07	.38	36	.19	.14	70	.42	.30
02	.23	.27	37	.46	.26	71	.48	.28
03	.11	.02	38	.23	.47	72	.07	.49
04	.34	.34	39	.03	.38	73	.20	.07
05	.47	.19	40	.39	.41	74	.28	.13
06	.26	.42	41	.06	.01	75	.24	.23
07	.02	.03	42	.34	.43	76	.48	.06
						77	.10	.16
08	.26	.34	43	.31	.37	78	.31	.06
09	.00	.19	44	.36	.50			
10	.03	.21	45	.44	.12	79	.33	.31
11	.34	.29	46	.15	.18	80	.40	.12
12	.47	.42	47	.26	.30	81	.24	.19
13	.26	.05	48	.42	.21	82	.10	.01
14	.33	.21	49	.42	.13	83	.45	.21
			50	.21	.27	84	.07	.47
15	.35	.46	51	.23	.14	85	.21	.07
16	.38	.13				86	.44	.05
17	.02	.37	52	.09	.08	87	.08	.04
18	.16	.32	53	.29	.40			
19	.38	.50	54	.02	.27	88	.18	.13
20	.18	.12	55	.25	.48	89	.07	.39
21	.49	.36	56	.37	.28	90	.23	.17
			57	.45	.31	91	.23	.40
22	.38	.33	58	.42	.08	92	.47	.33
23	.04	.32	59	.11	.36	93	.11	.34
24	.44	.14	60	.07	.05	94	.45	.13
25	.22	.38				95	.43	.24
26	.24	.12	61	.14	.00	96	.25	.30
27	.14	.18	62	.21	.01			
28	.08	.24	63	.35	.47			
			64	.12	.09			
29	.45	.45	65	.16	.44			
30	.03	.45	66	.33	.08			
31	.25	.26	67	.34	.19			
32	.16	.49	68	.19	.25			
33	.25	.13	69	.07	.29			
34	.05	.47						
35	.04	.25						



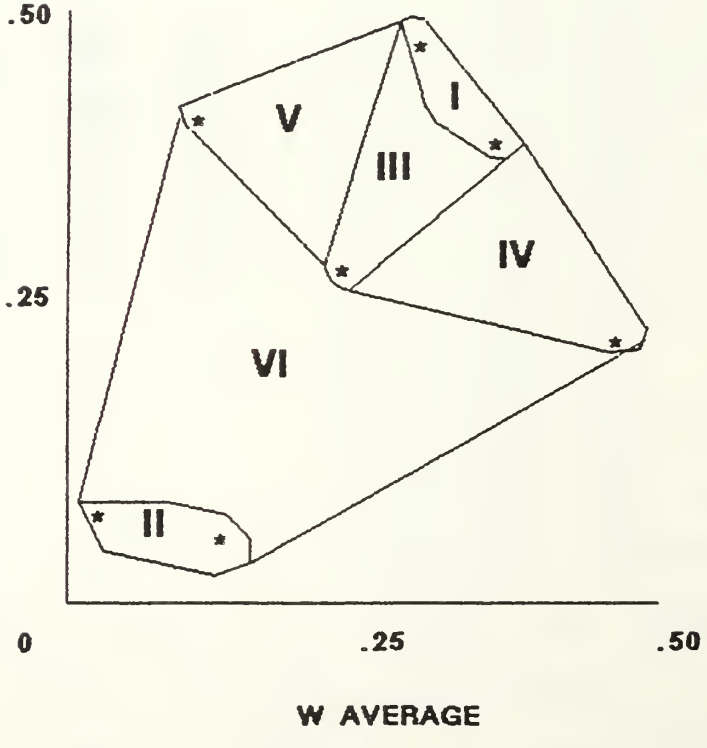


Figure E.1 Test Similarity--Example Data Set 1

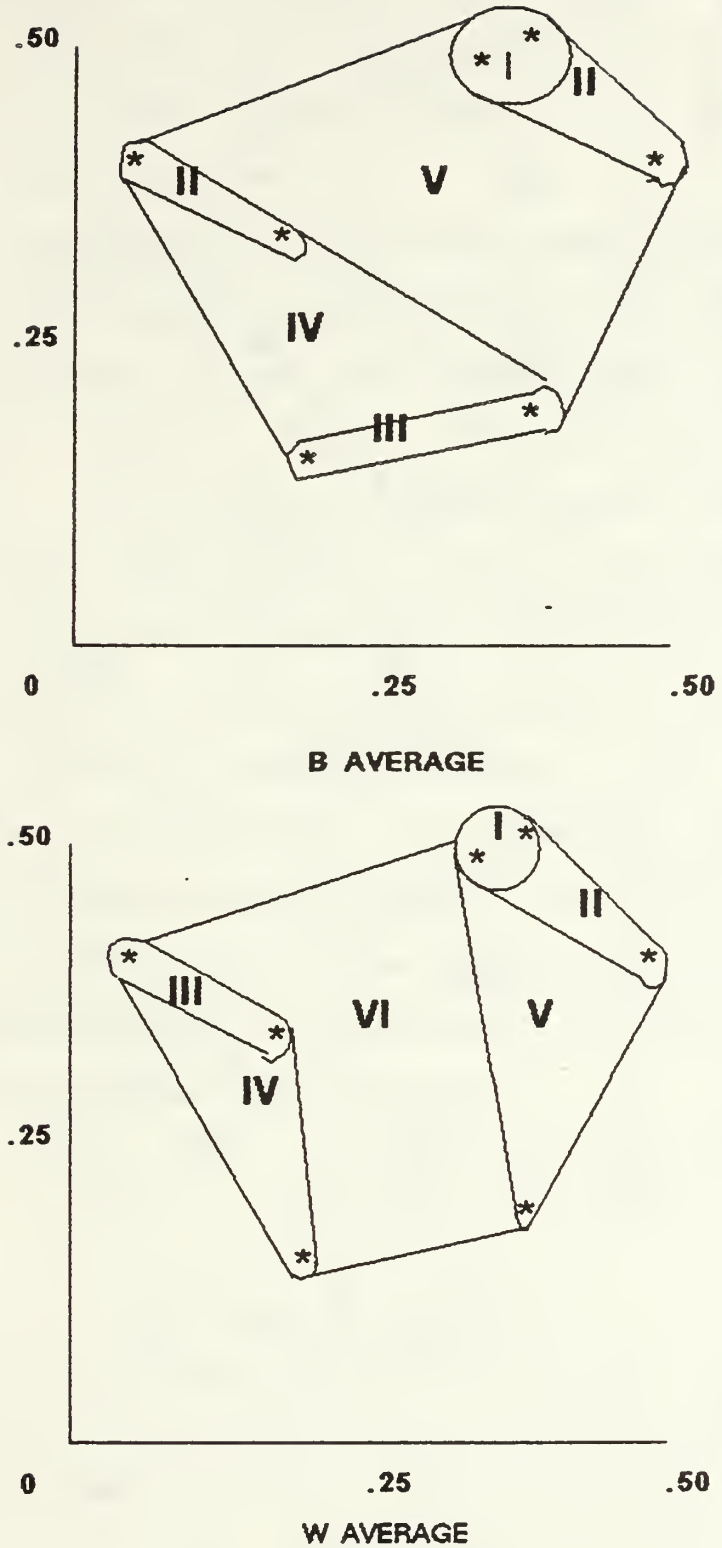
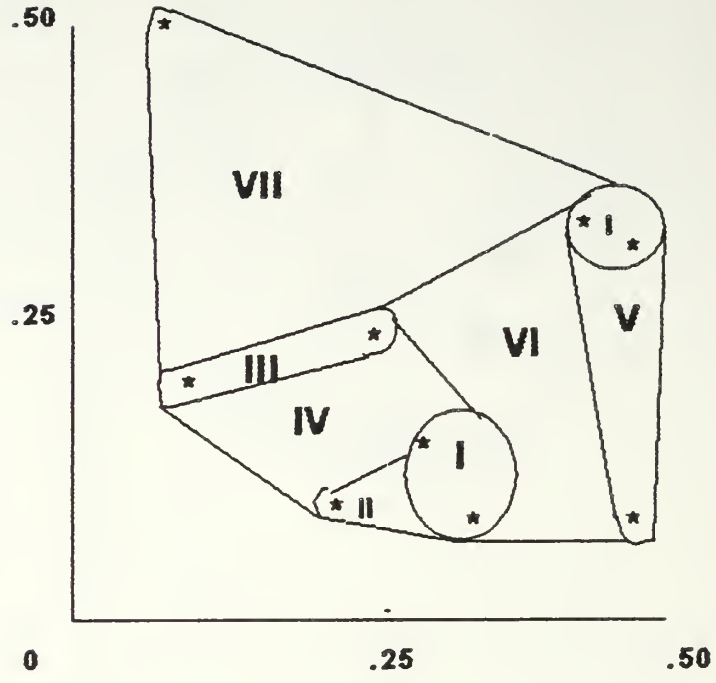


Figure E.2 Test Dissimilarity--Example Data Set 3



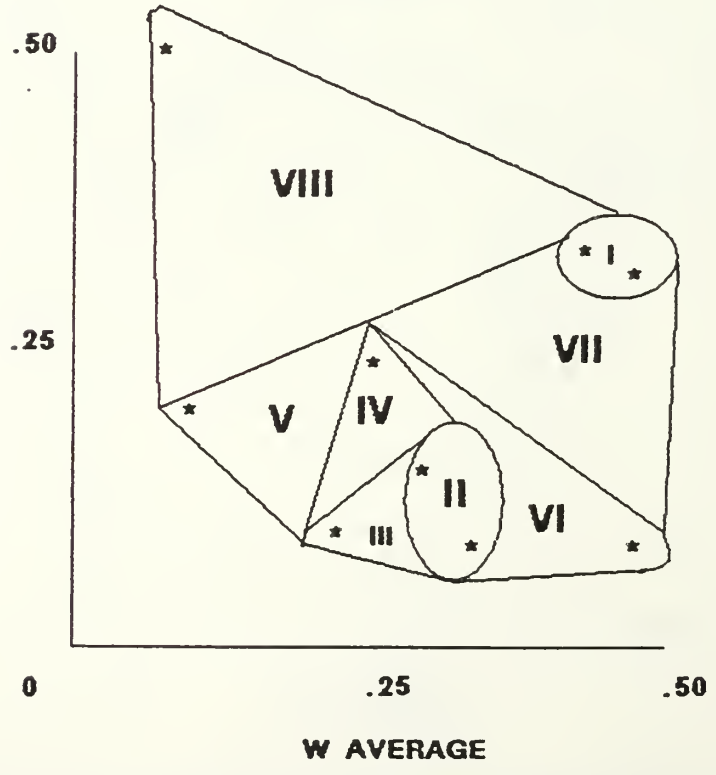


Figure E.3 Test Dissimilarity--Example Data Set 10

The Roman numerals identify the sequence of clustering. Often the clustering of cases occur at the same level (distance) as illustrated in Figure E.2, B AVERAGE, where two clusters are sequenced II.

In the tests which demonstrated interestingly different clustering results, the average linkage within groups method tended to cluster one or two distinct sets initially and expand the clusters quickly into higher levels of aggregation. The average linkage between groups method, however, created more clusters initially and pooled clusters into higher degrees of aggregation later in the sequence.

More clusters at the lowest level of agglomerative hierarchy provide greater insight into data set relationships characterized by inherently small ratio differences. The tendency to create more clusters from data initially is agreeable with the needs of this project. The average linkage between groups method was therefore selected as the preferred clustering criterion.

APPENDIX F
CLUSTER STRENGTH TABLE PROGRAM

```
MCDEN:PROC OPTIONS(MAIN);                                                    MCD00010
 DCL LINE CHAR(133) VARYING,                                                 MCD00020
     1 STREC,    /* STRENGTH RECORD */                                       MCD00030
       2 MOS BIN FIXED(15),                                                  MCD00040
       2 INV(11) BIN FIXED(31),                                              MCD00050
                                                                             MCD00060
     ST FILE INPUT STREAM ENV(F(115)),                                       MCD00070
     CL FILE INPUT STREAM,                                                   MCD00080
     FON BIT(1),                                                             MCD00090
     XMOS BIN FIXED(15),                                                     MCD00100
     TEMP CHAR(100) VARYING,                                                 MCD00110
     VMOS(200) BIN FIXED(15),                                                MCD00120
     VSTRN(200) BIN FIXED(31),                                               MCD00130
     (IVPTR,ILINE,IP,IH,IL,NMOS) BIN FIXED(15),                              MCD00140
     IM BIN FIXED(31),                                                       MCD00150
     CLUSTER(23,23) BIN FIXED(31);                                           MCD00160
                                                                             MCD00170
                                                                             MCD00180
 ON ENDFILE(ST) FON='0'B;                                                    MCD00190
 ON ENDFILE(CL) FON='0'B;                                                    MCD00200
 OPEN FILE(CL) TITLE('CL');                                                  MCD00210
                                                                             MCD00220
 NMOS=0;                                                                     MCD00230
 FON='1'B;                                                                   MCD00240
 CALL GETYR;    /* GET YR POINTER */                                         MCD00250
 GET FILE(ST) EDIT(STREC) (F(5),11 F(8));                                    MCD00260
 DO WHILE(FON);                                                              MCD00270
  NMOS=NMOS+1;                                                               MCD00280
  VMOS(NMOS)=MOS;                                                            MCD00290
  VSTRN(NMOS)=INV(IVPTR);                                                    MCD00300
  GET FILE(ST) SKIP EDIT(STREC) (F(5),11 F(8));                              MCD00310
 END;                                                                        MCD00320
                                                                             MCD00330
 FON='1'B;                                                                   MCD00340
 DO WHILE(FON);                                                              MCD00350
  CALL GETSYN;  /* GET 1ST LINE OF CLUSTER MEMBERSHIP */                     MCD00360
  DO WHILE(FON & ILINE<=44 & SUBSTR(LINE,40,20)      =' ' &                       MCD0037
                                        INDEX(LINE,'CLUSTER')=0);MCD00380
   CALL GETCL;   /* ACCUMULATE CLUSTER INFO */                               MCD00390
   GET FILE(CL) EDIT(LINE) (A(133));                                         MCD00400
   ILINE=ILINE+1;                                                            MCD00410
  END;                                                                       MCD00420
END;                                                                         MCD00430
                                                                             MCD00440
 PUT PAGE EDIT( (I DO I=IH TO IL BY -1) ) (X(5),22 F(5));                    MCD00450
 PUT SKIP(1);                                                                MCD00460
 DO I=1 TO IH;                                                               MCD00470
  PUT SKIP EDIT(I,(CLUSTER(I,J) DO J=IH TO MAX(I,IL) BY -1))(23 F(5));       MCD00480
 END;                                                                        MCD00490
                                                                             MCD00500
GETYR:PROC;     /* GET YR SELECTED */                                        MCD00510
 LINE=' ';                                                                   MCD00520
 DO WHILE(FON & INDEX(LINE,'SELECT IF (YR EQ')=0);                           MCD00530
  GET FILE(CL) EDIT(LINE) (A(133));                                          MCD00540
 END;                                                                        MCD00550
 I=INDEX(LINE,'YR EQ');                                                      MCD00560
 I=SUBSTR(LINE,I+6,2);                                                       MCD00570
```

```
                                                                        MCD00580
  IF I=98 THEN IVPTR=10;                                                MCD00590
   ELSE IF I=99 THEN IVPTR=11;                                          MCD00600
   ELSE IVPTR=I-77;                                                     MCD00610
                                                                        MCD00620
  PUT SKIP LIST('**YEAR SELECTED=',I,IVPTR);                            MCD00630
  IF IVPTR<1 | IVPTR>11 THEN                                            MCD00640
        DO; PUT SKIP LIST('***ERROR: SELECT YR RECORD NOT FOUND');      MCD00650
            STOP;                                                       MCD00660
        END;                                                            MCD00670
END GETYR;                                                              MCD00680
                                                                        MCD00690
GETSYN:PROC;                                                            MCD00700
 DO WHILE(FON & INDEX(LINE,'CLUSTER MEMBERSHIP')=0);                    MCD00710
  GET FILE(CL) EDIT(LINE) (A(133));                                     MCD00720
 END;                                                                   MCD00730
 IF  FON THEN RETURN;                                                   MCD00740
                                                                        MCD00750
 GET FILE(CL) EDIT(LINE) (SKIP(4),A(133));                              MCD00760
 IH=SUBSTR(LINE,17,5);                                                  MCD00770
 IL=IH;                                                                 MCD00780
 IP=17;                                                                 MCD00790
 DO I=IH-1 TO 2 BY -1;                                                  MCD00800
  IP=IP+5;                                                              MCD00810
  TEMP=SUBSTR(LINE,IP,5);                                               MCD00820
  IF TEMP  =' ' THEN IL=TEMP;                                           MCD00830
 END;                                                                   MCD00840
 GET FILE(CL) EDIT(LINE) (SKIP(2),A(133));                              MCD00850
 ILINE=1;                                                               MCD00860
       PUT SKIP LIST('=== IH,IL=',IH,IL);                               MCD00870
END GETSYN;                                                             MCD00880
                                                                        MCD00890
GETCL:PROC;   /* ACCUMULATE CLUSTER INFO */                             MCD00900
 XMOS=SUBSTR(LINE,6,3);                                                 MCD00910
 IM=GETMOS(XMOS);   /* GET INVENTORY STRENGTH FOR MOS */                MCD00920
 IP=12;                                                                 MCD00930
 DO I=IH TO IL BY -1;                                                   MCD00940
  IP=IP+5;                                                              MCD00950
  IV=SUBSTR(LINE,IP,5);   /* CLUSTER NO. WHERE MOS BELONGS */           MCD00960
  IF IV>0 THEN CLUSTER(IV,I)=CLUSTER(IV,I) + IM;                        MCD00970
 END;                                                                   MCD00980
END GETCL;                                                              MCD00990
                                                                        MCD01000
GETMOS:PROC(XMOS) RETURNS(BIN FIXED(31)); /* GET INVENTORY */           MCD01010
 DCL XMOS BIN FIXED(15),                                                MCD01020
     I BIN FIXED(15);                                                   MCD01030
 DO I=1 TO NMOS;                                                        MCD01040
  IF XMOS=VMOS(I) THEN RETURN(VSTRN(I));                                MCD01050
 END;                                                                   MCD01060
 PUT SKIP LIST('*** ERROR.  MOS NOT FOUND IN VMOS ',XMOS);              MCD01070
 STOP;                                                                  MCD01080
END GETMOS;                                                             MCD01090
                                                                        MCD01100
END MCDEN;                                                              MCD01110
```

LIST OF REFERENCES

1. Tucker, D.D., Loss Rate Estimation in Marine Corps Officer Manpower Models, Master's Thesis, Naval Postgraduate School, Monterey, California, September 1985.

2. Robinson, J.R., Limited Translation Shrinkage Estimation of Loss Rates in Marine Corps Manpower Models, Master's Thesis, Naval Postgraduate School, Monterey, California, March 1986.

3. Butterworth, R.W. and Milch, P.R., "Clustering Navy Ratings by Loss Behavior," Naval Postgraduate School, Monterey, California, 1975.

4. Decision System Associates, Inc., "Functional Description for the Development of the Officer Planning and Utilization System (OPUS)," Rockville, Maryland, 1986.

5. Navy Personnel Research and Development Center, "System Design for the Marine Corps Officer Rate Projector (MCORP)," San Diego, California, 1985.

6. Elseramegy, A., CART Program: The Implementation of the Classification and Regression Tree Resubstitution Implementation Application, Master's Thesis, Naval Postgraduate School, Monterey, California, December 1985.

7. Siegel, B., "Methods for Forecasting Officer Loss Rates," Navy Personnel Research and Development Center, San Diego, California, 1983.

8. Bres and Rowe, "Development and Analysis of Loss Rate Forecasting Technique for the Navy's Unrestricted Line Officers," Navy Personnel Research and Development Center, San Diego, California, 1979.

9. Decision System Associates, Inc., "User's Manual for the Officer Rate Generator," Rockville, Maryland, 1985.

10. Decision System Associates, Inc., "OPUS--System Specification," Rockville, Maryland, 1986.

11. Decision Systems Associates, Inc., "OPUS--System Specifications for Optimum Officer Force Model," Rockville, Maryland, 1987.

12. Decision Systems Associates, Inc., "OPUS--System Specifications for Officer Population Simulation," Rockville, Maryland 1986.

13. Decision Systems Asscciates, Inc., "User's Manual for Officer Planning and Utility System (OPUS)," Rockville, Maryland, 1985.

14. Bartholomew, D.J., and Forbes, A.F., Statistical Techniques for Manpower Planning, Wiley, 1979.

15. Berenson, M.L., Levine, and Goldstein, Intermediate Statistical Methods and Applications, Prentice Hall, 1983.

16. Grinold, R.C. and Marshall, K.T., Manpower Planning Models, North Holland, 1977.

17. Johnson, S.C., "Hierarchical Clustering Schemes," Psychometrika, V. 32, pp. 241-254, September 1967.

18. Anderberg, M.R., Cluster Analysis for Applications, Academic Press, 1973.

19. Lorr, M., Cluster Analysis for Social Scientists, Jossey-Bass Publishers, 1983.

20. Norusis, M.J., Advanced Statistics Guide--SPSSX, McGraw-Hill Book Company, 1985.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Professor Robert R. Read, Code 55Re Naval Postgraduate School Monterey, California 93943-5004	5
4.	Professor George W. Thomas, Code 54Te Naval Postgraduate School Monterey, California 93943-5004	2
5.	Commandant of the Marine Corps HQMC, Code MPP-30 Washington, D.C. 22314	10
6.	Commanding Officer Navy Personnel Research and Development Center San Diego, California 92152	2
7.	Professor Richard L. Hall 410 N. Granada St. Arlington, Virginia 22203	1
8.	Captain Randall W. Larsen 15820 Lazy Day Lane Montclair Country Club Dumfries, Virginia 22026	3

Thesis
L27265 Larsen
c.1 The aggregation of population groups to improve the predictability of Marine Corps officer attrition estimation.